AD_____

AWARD NUMBER: W81XWH-04-1-0475

TITLE: Development of a Computer-Aided Diagnosis System for Early Detection of Masses Using Retrospectively Detected Cancers on Prior Mammograms

PRINCIPAL INVESTIGATOR: Jun Wei, Ph.D.

CONTRACTING ORGANIZATION: University of Michigan
Ann Arbor, Michigan 48109-1274

REPORT DATE: June2006

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
01-06-2006	Annual	1 Jun 2005 – 31 May 2006

4. TITLE AND SUBTITLE

Development of a Computer-Aided Diagnosis System for Early Detection of Masses Using Retrospectively Detected Cancers on Prior Mammograms

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-04-1-0475

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Jun Wei, Ph.D.

E-Mail: jvwei@umich.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Michigan
Ann Arbor, Michigan 48109-1274

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The goal of this project is to develop a computer-aided diagnosis (CAD) system for mass detection using advanced computer vision techniques that will be trained with retrospectively detected cancers on prior mammograms. The new CAD system will be combined with our existing CAD system. When fully developed, the new dual CAD system should increase the sensitivity of detecting cancers at the early stage without compromising the sensitivity for other cancers. During this project year, we have performed the following tasks: (1) continue to collect the data sets of digitized film mammograms with multiple examinations, (2) investigate the performance of a regularized discriminant analysis (RDA) classifier in combination with a feature selection method for classification of masses and normal tissues on mammograms, (3) develop a two-view information fusion method to improve the performance of our CAD system, and (4) continue to develop a fusion scheme for combining two CAD systems to improve detection of early stage breast cancer. In summary, we have investigated a number of areas in CAD of mammographic masses and evaluated the new techniques for mass detection on mammograms. We have made progress in three of the tasks proposed in the project. We have found that our new computer-vision techniques can improve the performance of the CAD systems. We will continue the development of the CAD system in the coming years.

15. SUBJECT TERMS
No subject terms provided.

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			USAMRMC
U	U	U	UU	93	**19b. TELEPHONE NUMBER** *(include area code)*

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

(3) Table of Contents

1. Introduction

Recent clinical studies have proved that computer-aided diagnosis (CAD) systems are helpful for improving cancer detection by radiologists on mammograms[1-6]. To evaluate the effectiveness of a CAD system in detecting cancers that are likely to be missed by radiologists, one way is to study its accuracy in detecting missed cancers on prior mammograms (the mammograms in previous exams on which the cancer can be seen retrospectively). Several studies have demonstrated that CAD systems have potential ability to detect missed cancers on prior mammograms[7-11]. However, the performance of a CAD system on prior mammograms is generally much lower than their performance on the current mammograms (the mammogram on which cancer is detected). Recently, one study investigated the performance change between prior mammograms and current mammograms when using the CAD system trained by current mammograms and another by prior mammograms. It was concluded that CAD schemes trained with the current mammograms do not perform optimally in detecting masses depicted on prior images and vice versa.

The goal of this proposed project is to develop a CAD system using advanced computer vision techniques to detect masses using retrospectively detected cancers on prior mammograms and incorporate the developed CAD system into our current CAD system. We hypothesize that a dual CAD system, which combines a system trained with subtle lesions retrospectively seen on prior mammograms and a system trained with cancers detected on current mammograms, should increase the sensitivity of detecting cancers at the early stage without compromising its ability to detect less subtle cancers. To accomplish this goal, we will (1) collect a large database of masses on digitized prior and current film mammograms (DFMs) for training and testing the CAD system, (2) develop single-view computer vision techniques for mass detection and classification in prior DFMs, (3) reduce false positives (FPs) by correlation of image information from two-view mammograms, (4) combine the new CAD system with our current CAD system without an increase in overall FPs, and (5) perform ROC study to evaluate the effects of CAD on radiologists' accuracy in detecting subtle cancers. Although we do not plan to develop such a system for digital mammograms because there will not be enough prior digital mammograms with cancers available for the development, the general methodology developed in this study can be adapted to CAD systems for digital mammograms in the future.

At the conclusion of this project, we expect that a fully automated CAD system will be developed which can be used for detection of masses on DFMs. The general methodology developed in this study may also be adapted to develop similar software for other CAD systems. The significance of this project is that it will develop a CAD system which can further improve radiologists' accuracy in detecting breast cancers at an early stage. Since early detection and treatment can reduce breast cancer mortality rate, the CAD system will be useful for increasing the effectiveness of mammographic screening.

(5) Body

The current year (6/1/05-5/31/06) is the second year of the project. We will describe in the following details of the studies that we performed this year.

(A) Collection of a Database of Digitized Screen-film Mammograms (DFM) with Multiple Examinations

In this project year, we continue to collect a data set of digitized screen-film mammogram from patient files in the Department of Radiology at the University of Michigan with Institutional Review Board (IRB) approval. Two independent data sets of mammograms were collected for this study; one contained mammograms with masses and the other contained normal mammograms. The normal data set was used to estimate the false positive (FP) marker rates during testing[12-14]. To date, the mass data set contained 160 cases with 160 masses. 90 of the masses are biopsy proven to be malignant and 56 to be benign. The remaining 14 masses are considered benign by long-term follow-up. Each case included the current mammograms on which the mass was detected by radiologists, and the prior mammograms obtained from previous exams. The mass set contained 320 current mammograms and 406 prior mammograms. The true location of each mass was identified by an experienced Mammography Quality Standards Act (MQSA) radiologist. The radiologist also measured the mass size and provided descriptions of the mass margin, shape, conspicuity, and breast density.

(B) Investigation of a Regularized discriminant analysis for breast mass detection

The first study of this project is to develop a single CAD system for mass detection on prior DFMs. In computer-aided detection (CAD) applications, an important step is to design a classifier for the differentiation of the abnormal from the normal structures. We have previously developed a stepwise linear discriminant analysis (LDA) method with simplex optimization for this purpose. In this year, we have performed a preliminary study to investigate the performance of a regularized discriminant analysis (RDA) classifier in combination with a feature selection method for classification of the masses and normal tissues detected on mammograms. Our preliminary results were presented at the SPIE meeting in 2006[15]. The study is summarized in the following.

1) Data Set

IRB approval was obtained prior to the commencement of this investigation. The images used in this study were acquired at the University of Michigan with a GE Senographe 2000D FFDM system before biopsy. The GE system has a CsI phosphor/a:Si active matrix flat panel digital detector with a pixel size of $100\mu m \times 100\mu m$ and 14 bits per pixel. A data set of 130 cases was used. All cases had two mammographic views, the craniocaudal (CC) view and the mediolateral oblique (MLO) view or the lateral (LM or ML) view. The data set contained 130 biopsy-proven masses. The true locations of the masses were identified by a Mammography Quality Standards Act radiologist.

2) Methods

2.1) Discriminant Analysis

Assume that the class distributions are multivariate normal in a two-class classification problem. Under this condition, discriminant analysis models differ essentially by the specific assumptions on the mean vectors and covariance matrices of the group conditional densities. The most commonly used model is linear discriminant analysis (LDA) which assumes that the group conditional distributions are multivariate normal distributions with mean vectors μ_k, where $k = 1, 2$ is the class index, and equal covariance matrix Σ. The definition of LDA is given in Eq. (1).

$$Y = (\mu_1 - \mu_2)^T \Sigma^{-1} X \tag{1}$$

where $X^T=(x_1, \ldots, x_n)$ is the feature vector of a sample and n is the dimensionality of the feature space. If the covariance matrices are not equal, one can use quadratic discriminant analysis (QDA), which has a quadratic term for the feature vector in its model. The definition of QDA is described in Eq. (2).

$$Y = \frac{1}{2} X^T (\Sigma_2^{-1} - \Sigma_1^{-1}) X - (\mu_2^T \Sigma_2^{-1} - \mu_1^T \Sigma_1^{-1}) X \tag{2}$$

The parameters in LDA and QDA are usually unknown and have to be estimated from training samples. In medical imaging applications, the sample size may be very small in comparison with the dimensionality of the feature space. A regularization technique for discriminant analysis, referred to as regularized discriminant analysis (RDA)[16], makes use of a complexity parameter and a shrinkage parameter to design an intermediate classification model between LDA and QDA. The covariance matrices can thus be written as:

$$\hat{\Sigma}_k = (1-\gamma)\Sigma_k + \frac{\gamma}{p} tr[\Sigma_k] I \quad , \quad k=1, 2 \tag{3}$$

where I is the identity matrix, γ and p are the complexity parameter and the shrinkage parameter, respectively. In this work, we investigated the use of the RDA classifier for FP reduction in a mass CAD system.

2.2) Feature Selection

In order to obtain the best texture feature subset and reduce the dimensionality of the feature space to design an effective classifier, feature selection was applied to the training set. Stepwise LDA feature selection with Wilks' lambda as the selection criterion was employed in our previous study. Simplex optimization procedure was used to choose the best set of feature selection parameters which includes a threshold F_{in} for feature entry, a threshold F_{out} for feature removal, and a tolerance threshold T for excluding features that have high correlation with the features already in the selected pool. In this study, we compared a new stepwise feature selection procedure with the current method. In the proposed method, a feature selection scheme which combines forward stepwise feature selection and backward stepwise feature elimination is used to obtain the best feature subset, using the area under the receiver operating characteristic (ROC) curve, A_z, as the selection criterion instead of Wilks' lambda. We evaluated the classifier performance using a leave-one-case-out resampling scheme within the training set, the test

discriminant scores from the left-out cases were analyzed using ROC methodology. The discriminant scores were input as the decision variable in the LABROC program, which fits a binormal ROC curve based on maximum likelihood estimation. The performances of the RDA classifier and the LDA classifier, both with the new feature selection method, were compared to that of the LDA classifier using the Wilks' lambda as the stepwise feature selection criterion in terms of their A_z for the classification of masses and normal tissue.

3) Results

We randomly separated the cases in our data set into two independent data subsets: 66 and 64 cases. The training and testing were performed using the cross validation method. The detection performance of the CAD system was assessed by free response receiver operating characteristic (FROC) analysis. FROC curves were presented on a per-mammogram and a per-case basis. For mammogram-based FROC analysis, the mass on each mammogram was considered as an independent true object. For case-based FROC analysis, the same mass imaged on the two-view mammograms was considered to be one true object and the detection of either or both masses on the two views was considered to be a true-positive (TP). The average test FROC curve was obtained by averaging the FP rates at the same sensitivity along the two corresponding test FROC curves from the 2-fold cross validation. The CAD system using RDA with the new feature selection method achieved an image-based sensitivity of 60%, 65%, and 70% at 1.1, 1.4, and 1.6 FPs/image, respectively, compared with 1.4, 1.7, and 2.1 FPs/image for the CAD system using LDA with the new feature selection method. The CAD system with stepwise LDA and simplex optimization achieved FP rates of 1.6, 1.9, and 2.2 FPs/image, respectively, at the same sensitivities, which were comparable to the FP rates of the CAD system using LDA with the new feature selection method. Figures 1(a) and 1(b) show the comparison of the image-based and case-based average FROC curves of the CAD systems using the three different classification methods, respectively.

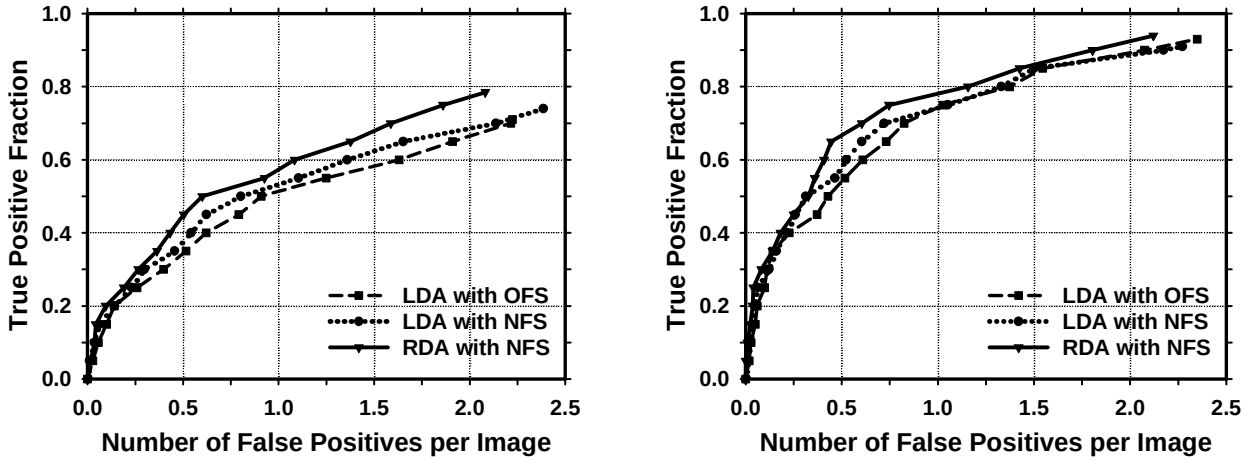


Figure 1. Comparison of FROC curves. OFS: stepwise feature selection with simplex optimization. NFS: feature selection combining forward feature selection and backward feature elimination. (a) image-based FROC curve, (b) case-based FROC curve.

(C) Development of a two-view information fusion method

The second study performed in this project year is to develop a two-view information fusion method to improve the performance of our CAD system for mass detection. Our preliminary results were presented at the SPIE meeting in 2006[17]. The study is summarized in the following.

1) Data Set

All mammograms in this study were collected from patient files at the University of Michigan with IRB approval. The mammograms were digitized with a LUMISYS 85 laser film scanner with a pixel size of 50μm×50μm and 4096 gray levels. The scanner was calibrated to have a linear relationship between gray levels and optical densities (O.D.) from 0.1 to greater than 3 O.D. units. The nominal O.D. range of the scanner is 0–4. The full resolution mammograms were first smoothed with a 2×2 box filter and subsampled by a factor of 2, resulting in images with a pixel size of 100μm×100μm. These images were used for the input of our CAD system. The data set we used in this study contained 475 cases, of which 464 cases had the two-view mammograms (the CC view and the MLO view or the lateral view) and 11 cases had four-view mammograms, resulting in a total of 972 mammograms. All mammograms were obtained before biopsy. There were 475 biopsy-proven masses in this data set.

2) Methods

In order to improve the overall performance of our CAD system for detection of masses, we developed a two-view fusion technique which combines the information from two mammographic views. The fusion method used in this study is based on the assumption that the corresponding true mass on two different mammographic views will exhibit similarities in their geometric, morphological and textural features which are relatively invariant with respect to the imaging views. On the other hand, FPs detected by CAD system are expected to exhibit a lesser degree of similarity because they are usually objects formed by different normal tissues.

For a given object on one view, geometric pairing is first performed using the nipple-to-object distance as the average radius of an annular region on the other view within which the detected objects can be paired with the given object. Manually identified nipple locations are used for the registration in this study. We are developing an automated nipple detection technique[18] and the automated method will be used when it reaches high accuracy. Similarity measures between each pair of objects are derived from the pairs of individual object features. The similarity features include morphological features, Hessian feature, correlation coefficients between the two paired objects and texture features. A similarity classifier is trained to distinguish between true and false pairs by merging the similarity features into a similarity score for each object. The similarity score and the single-view object score of the object are then fused to form a final score for the object. Our two-view system is summarized in Figure 2.

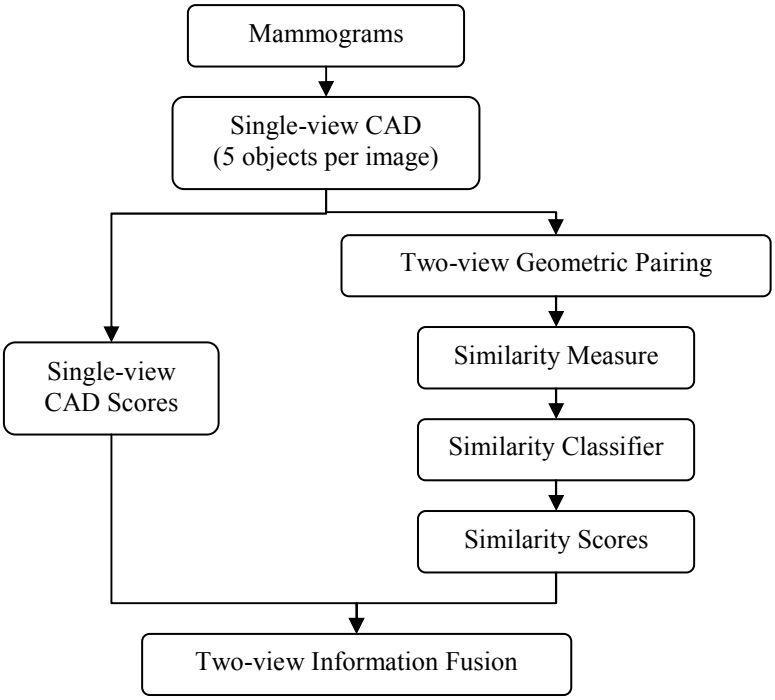


Figure 2. Block diagram of the two-view CAD system for mass detection on mammograms.

3) Results

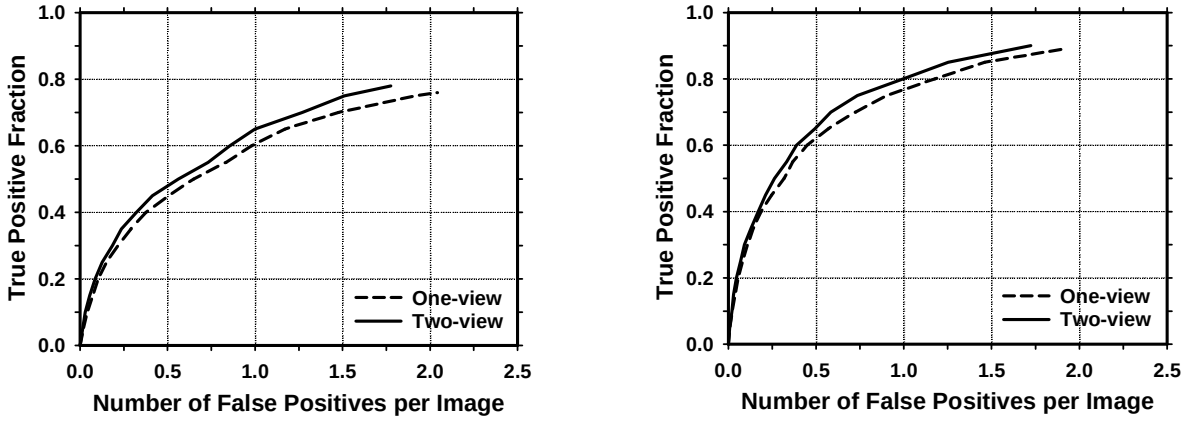


Figure 3. Comparison of the average test FROC curves obtained by averaging the FROC curves from the two independent mass subsets. Single-view: detection by the single-view CAD system. Two-view: detection by the two-view CAD system. (a) Image-based FROC curves, (b) Case-based FROC curves.

We randomly separated the cases in our data set into two independent equal sized data sets: 243 cases with 494 images and 232 cases with 478 images. The training and testing were performed using the 2-fold cross validation method. The detection performance of the CAD system was assessed by FROC analysis. FROC curves were presented on a per-mammogram and a per-case basis. To evaluate the overall test performance, an average test FROC curve was obtained as described above. When the single-view CAD system was applied to the test set, the FPs/image were 2.0, 1.5, and 1.2 at the case-based sensitivities of 90%, 85% and 80%, respectively. With the two-view CAD system, the FP rates were improved to 1.7, 1.3, and 1.0 FPs/image at the same case-based sensitivities. Figures 3(a) and 3(b) show the comparison of the test performance of the single-view CAD system and the two-view CAD systems by using image-based and case-based average FROC curves, respectively.

(D) Development of a fusion scheme to combine two CAD systems

In this project year, we continued to develop a fusion scheme to combine two single CAD systems. We have recently submitted a journal paper to Medical Physics[19]. The detailed methods and results of the study can be found in the enclosed manuscript (Appendix). The study is summarized in the following.

1) Data Set

We collected three data sets. The first data set contained 115 cases with confirmed masses. Each case included the current mammograms that prompted the radiologist to work up the mass. This is referred to as the "average" mass set. All of the cases in the average mass set had two mammographic views: the CC view and the MLO view or the lateral view, thus yielding a total of 230 mammograms. There were 115 masses (67 malignant masses and 48 benign masses) in this data set, of which 105 were biopsy-proven and 10 were determined to be benign by long-term follow-up.

The second data set was composed of the prior mammograms dated one to two years earlier than the mammograms of the same patients in the average mass set. Since the masses on prior mammograms are on average subtler than those on current mammograms, this data set is referred to as the "subtle" mass set. On five of the 115 patients, no mass or focal density could be identified on either view of the prior mammograms. Therefore, the subtle mass set was composed of 110 cases (62 malignant and 48 benign). For the purpose of training the subtle mass detection system, the subtle masses do not have to be obtained from the same cases as the average mass set but we used the available prior mammograms for these mass cases in our database. Nineteen of the 110 cases had two prior mammogram examinations. Of the 129 examinations in the subtle mass set, 123 had two mammographic views and 6 had three views, with a total of 264 mammograms.

The third data set was composed of 260 normal bilateral two-view mammograms obtained from 65 patients. No masses were evident on these mammograms upon review by the experienced radiologist.

2) Methods

During the training of the dual system, we used the current and prior mammograms from the same patients. The current mammograms that contained the average masses were only used to train the first single CAD system. The prior mammograms that contained the subtle masses were only used to train the second single CAD system. The prescreening and the segmentation steps in the two systems are identical. Since the morphological appearances of average and subtle masses are different, the rules in the morphological rule-based FP classification are trained differently for the two single CAD systems. During testing with an independent mammogram, the dual system keeps all the suspicious objects that satisfy the FP classification rules of either single CAD system and applies the LDA classifiers from both single systems to each object. Each object thus has two LDA scores.

To merge the information from the two CAD systems, a fusion scheme was developed for our dual system. In this study, a feed-forward backpropagation artificial neural network (BP-ANN) was trained to classify the masses from normal tissues by combining the output information from the two single CAD systems. The LDA classifiers from the two single CAD systems were applied to each detected object. The two LDA discriminant scores for each object were used as input to the BP-ANN. The BP-ANN had an input layer with two nodes, a hidden layer with N nodes, and an output layer with one node. The nodes were interconnected by weights and information propagated from one layer to the next through a log-sigmoidal activation function. The learning of the ANN was a supervised process in which known training cases were input to the ANN. The performance function for the network was the mean-squared error between the network outputs and the target outputs. The weights of the network were adjusted iteratively by a feedforward backpropagation procedure to minimize the error. Detailed description of the backpropagation neural network can be found in the literature[20,21].

To test the dual system, the two trained single CAD systems, one trained with the average mass set and the other with the subtle mass set, were applied in parallel to each single "unknown" mammogram in the independent test subset. No prior mammogram was needed during testing.

3) Results

An important purpose of a CAD system is to serve as a second reader to alert radiologists to subtle cancers that may be overlooked. In this project year, we compared the performance of dual system approach with single CAD system using the data set with subtle masses in prior mammograms. Figure 4 and Figure 5 compare the average FROC curves of the single CAD system and the dual system for detection in the test subsets. The TP rate in Figure 4 was estimated by including both malignant and benign masses and that in Figure 5 was estimated from malignant masses only. The single CAD system trained with average masses alone was used. The FP rates for both systems were estimated from the mammograms without masses. The dual CAD system achieved a case-based sensitivity of 50% at 0.7 FP marks/image for all masses and at 0.5 FP marks/image for malignant masses only, compared with 1.4 FP marks/image for all

masses and 1.1 FP marks/image for malignant masses only using the single CAD system. The results in Fig. 5 indicate that if the CAD system threshold was set at about 0.5 FP marks/image, about 50% of the malignant masses in our database could have been detected and pointed out to the radiologists by the dual CAD system on the prior mammograms. The radiologists might have worked up some of these malignant masses during the prior exam and found these cancers earlier. Without the new dual CAD system approach, about 25% of the malignant cancers would still be detected by the use of a regular CAD system, but the benefit of CAD was almost doubled by the dual CAD system approach.

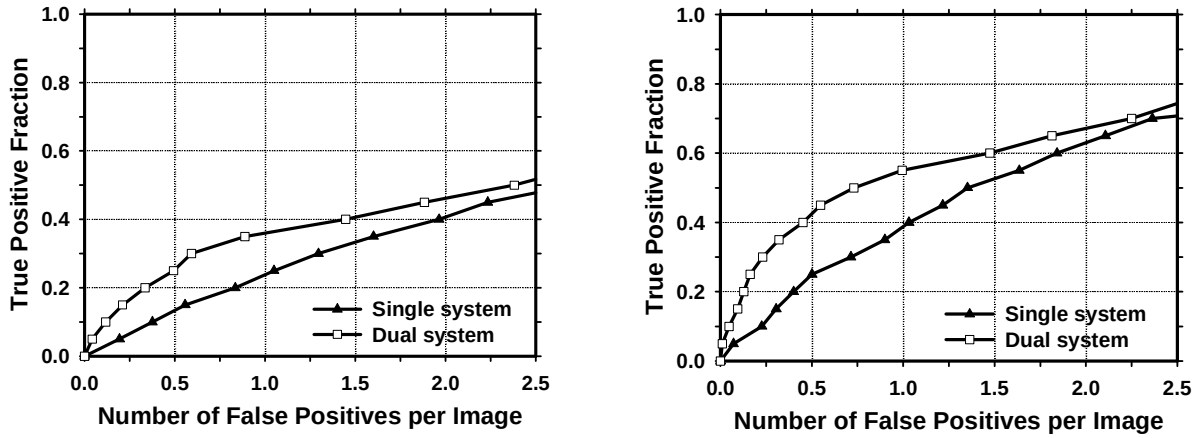


Figure 4. Comparison of the average test FROC curves for the single CAD system and the dual CAD system for detection of the subtle masses on the prior mammograms. The single CAD system trained with average masses alone was used and the FP rate was estimated from the mammograms without masses. (a) Image-based FROC curves, (b) Case-based FROC curves.

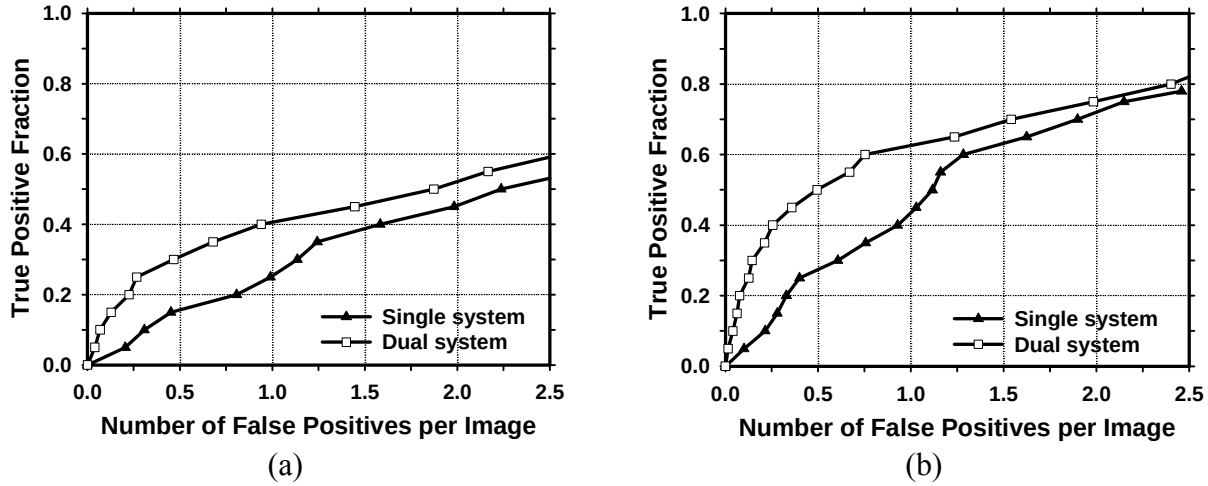


Figure 5. Comparison of the average test FROC curves for the single CAD system and the dual CAD system for detection of subtle malignant masses on the prior mammograms. The single CAD system trained with average masses alone was used and the FP rate was

estimated from the mammograms without masses. (a) Image-based FROC curves, (b) Case-based FROC curves.

(6) Key Research Accomplishments

- Continue to collect the data sets of digitized film mammograms with multiple examinations. (Task 1).
- Investigation of a Regularized discriminant analysis for breast mass detection (Task 2).
- Development of a two-view information fusion method (Task 3).
- Continue to develop a fusion scheme to combine two CAD systems (Task 4).

(7) Reportable Outcomes

As a result of the support by the USAMRMC BCRP grant, we have conducted studies to develop a computer-aided diagnosis system for early detection of masses using retrospectively detected cancers on prior mammograms. We have presented the results of these investigations in this project year and a journal article which was accepted for publication last year had been published in this project year. Also, we have submitted another journal paper to Medical Physics.

Journal Articles:

1. Wei J, Sahiner B, Hadjiiski LM, Chan HP, Petrick N, Helvie MA, Roubidoux MA, Ge J, Zhou C. Computer-aided detection of breast masses on full field digital mammograms. Medical Physics, Vol. 32, No. 9, pp. 2827-2837, 2005.
2. Wei J, Chan HP, Sahiner B, Hadjiiski LM, Helvie MA, Roubidoux MA, Zhou C, Ge J, "Dual system approach to computer-aided detection of breast masses on mammograms", Medical Physics (submitted)

Conference Proceeding:

1. Wei J, Sahiner B, Zhang Y, Chan HP, Hadjiiski LM, Zhou C, Ge J, and Wu YT, "Regularized discriminate analysis for breast mass detection on full field digital mammograms," SPIE Proc. 6144, 61445P-1~6 (2006).
2. Wei J, Sahiner B, Hadjiiski LM, Chan HP, Helvie MA, Roubidoux MA, Zhou C, Ge J, and Zhang Y, "Two-view information fusion for improvement of computer-aided detection (CAD) of breast masses on mammograms," SPIE Proc. 6144, 614424-1~7 (2006).

(8) Conclusions

During this project year, we first investigated the use of an RDA classifier with a new feature selection method to improve our single CAD system. Our results indicated that RDA in combination with the sequential forward inclusion-backward elimination feature selection method had potential to improve the performance of mass detection on mammograms. Further study is underway to test our method with a larger data set.

As a second study, we performed a preliminary study to develop a two-view information fusion method for improvement of our single CAD system. The improvement by using two-view information fusion was found to be statistically significant ($p<0.05$) by the AFROC method.

The third study in this project year is to continue to improve the performance of our mass detection system by using a new dual system approach which combines a CAD system optimized with "average" masses with another CAD system optimized with "subtle" masses. The statistical significance of the differences in the FROC curves of the different systems was estimated by using both the alternative free-response ROC (AFROC) method and the jackknife free-response ROC (JAFROC) method. Our results indicate that the dual CAD system approach can improve significantly ($p<0.05$) the performance of mass detection on mammograms compared to that obtained by training a single CAD system with the average masses alone or with both the average and the subtle masses by either the AFROC or the JAFROC method.

From the results of these studies, we found that our proposed dual CAD system approach is a very promising method to further improve radiologists' accuracy in detecting breast cancers at an early stage. We will continue to develop the CAD system in this direction in the coming project year.

(9) References

[1]T. W. Freer and M. J. Ulissey, "Screening mammography with computer-aided detection: Prospective study of 12,860 patients in a community breast center," Radiology 220, 781-786 (2001).

[2]M. A. Helvie, L. M. Hadjiiski, E. Makariou, H. P. Chan, N. Petrick, B. Sahiner, S. C. B. Lo, M. Freedman, D. Adler, J. Bailey, et al., "Sensitivity of noncommercial computer-aided detection system for mammographic breast cancer detection - A pilot clinical trial," Radiology 231, 208-214 (2004).

[3]R. L. Birdwell, P. Bandodkar, and D. M. Ikeda, "Computer-aided Detection with Screening Mammography in a University Hospital Setting," Radiology 236, 451-457 (2005).

[4]D. Gur, J. H. Sumkin, H. E. Rockette, M. A. Ganott, C. Hakim, L. A. Hardesty, W. R. Poller, R. Shah, and L. Wallace, "Changes in breast cancer detection and mammography recall rates after

the introduction of a computer-aided detection system," J National Cancer Institute 96, 185-190 (2004).

[5]S. A. Feig, E. A. Sickles, W. P. Evans, and M. N. Linver, "Re. Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system," J National Cancer Institute 96, 1260-1261 (2004).

[6]T. E. Cupples, "Impact of computer-aided detection (CAD) in a regional screening mammography program," Radiology 221(P), 520 (2001).

[7]N. Petrick, H. P. Chan, B. Sahiner, M. A. Helvie, and S. Paquerault, "Evaluation of an automated computer-aided diagnosis system for the detection of masses on prior mammograms," San Diego, 2000.

[8]G. M. Te Brake, N. Karssemeijer, and J. Hendriks, "Automated detection of breast carcinomas not detected in a screening program," Radiology 207, 465-471 (1998).

[9]D. M. Ikeda, R. L. Birdwell, K. F. O'Shaughnessy, E. A. Sickles, and R. J. Brenner, "Computer-aided detection output on 172 subtle findings on normal mammograms previously obtained in women with breast cancer detected at follow-up screening mammography," Radiology 230, 811-819 (2004).

[10]B. Zheng, W. F. Good, D. R. Armfield, C. Cohen, T. Hertzberg, J. H. Sumkin, and D. Gur, "Performance change of mammographic CAD schemes optimized with most-recent and prior image databases," Acad. Radiol. 10, 283-288 (2003).

[11]D. Gur, J. S. Stalder, L. A. Hardesty, B. Zheng, J. H. Sumkin, D. M. Chough, B. E. Shindel, and H. E. Rockette, "Computer-aided Detection Performance in Mammographic Examination of Masses: Assessment," Radiology 233, 418-423 (2004).

[12]K. F. O'Shaughnessy, R. A. Castellino, S. L. Muller, and K. Benali, "Computer-aided detection (CAD) on 90 biopsy-proven breast cancer cases acquired on a full-field digital mammography (FFDM) system," Radiology 221(P), 471 (2001).

[13]L. J. Warren Burhenne, S. A. Wood, C. J. D'Orsi, S. A. Feig, D. B. Kopans, K. F. O'Shaughnessy, E. A. Sickles, L. Tabar, C. J. Vyborny, and R. A. Castellino, "Potential contribution of computer-aided detection to the sensitivity of screening mammography," Radiology 215, 554-562 (2000).

[14]R. E. Brem, J. W. Hoffmeister, J. A. Rapelyea, G. Zisman, K. Mohtashemi, G. Jindal, M. P. DiSimio, and S. K. Rogers, "Impact of breast density on computer-aided detection for breast cancer," AJR Am J Roentgenol. 184, 439-444 (2005).

[15]J. Wei, B. Sahiner, Y. Zhang, H.-P. Chan, L. M. Hadjiiski, C. Zhou, J. Ge, and Y.-T. Wu, "Regularized discriminate analysis for breast mass detection on full field digital mammograms," SPIE Proc. 6144, 61445P-1~6 (2006).

[16]J. Friedman, "Regularized discriminant analysis," J. Amer. Stat. Assoc. 84, 165-175 (1989).

[17]J. Wei, B. Sahiner, L. M. Hadjiiski, H.-P. Chan, M. A. Helvie, M. A. Roubidoux, C. Zhou, J. Ge, and Y. Zhang, "Two-view information fusion for improvement of computer-aided detection (CAD) of breast masses on mammograms," SPIE Proc. 6144, 614424-1~7 (2006).

[18]C. Zhou, H. P. Chan, C. Paramagul, M. A. Roubidoux, B. Sahiner, L. M. Hadjiiski, and N. Petrick, "Computerized nipple identification for multiple image analysis in computer-aided diagnosis," Medical Physics 31, 2871-2882 (2004).

[19]J. Wei, H.-p. Chan, B. Sahiner, L. M. Hadjiiski, M. A. Helvie, M. A. Roubidoux, C. Zhou, and J. Ge, "Dual system approach to computer-aided detection of breast masses on mammograms," Medical Physics (2006 (submitted)).

[20]C. M. Bishop, *Neural Networks for Pattern Recognition*, (Clarendon Press, Oxford, 1995).

[21]J. A. Freeman and D. M. Skapura, *Neural Networks-Algorithms, Applications, and Programming Techniques*, (Addison-Wesley, Reading, 1991).

(10) Appendix

Copies of the following publications are enclosed with this report:

Journal Articles:

1. Wei J, Sahiner B, Hadjiiski LM, Chan HP, Petrick N, Helvie MA, Roubidoux MA, Ge J, Zhou C. Computer-aided detection of breast masses on full field digital mammograms. Medical Physics, 32, 2827-2837, 2005.

2. Wei J, Chan HP, Sahiner B, Hadjiiski LM, Helvie MA, Roubidoux MA, Zhou C, Ge J, "Dual system approach to computer-aided detection of breast masses on mammograms", Medical Physics (submitted)

Conference Proceedings:

1. Wei J, Sahiner B, Zhang Y, Chan HP, Hadjiiski LM, Zhou C, Ge J, and Wu YT, "Regularized discriminate analysis for breast mass detection on full field digital mammograms," SPIE Proc. 6144, 61445P-1~6 (2006).
2. Wei J, Sahiner B, Hadjiiski LM, Chan HP, Helvie MA, Roubidoux MA, Zhou C, Ge J, and Zhang Y, "Two-view information fusion for improvement of computer-aided

detection (CAD) of breast masses on mammograms," SPIE Proc. 6144, 614424-1~7 (2006).

Computer-aided detection of breast masses on full field digital mammograms

Jun Wei,[a] Berkman Sahiner, Lubomir M. Hadjiiski, Heang-Ping Chan, Nicholas Petrick, Mark A. Helvie, Marilyn A. Roubidoux, Jun Ge, and Chuan Zhou
Department of Radiology, University of Michigan, Ann Arbor, Michigan 48109

(Received 7 February 2005; revised 16 June 2005; accepted for publication 16 June 2005; published 23 August 2005)

We are developing a computer-aided detection (CAD) system for breast masses on full field digital mammographic (FFDM) images. To develop a CAD system that is independent of the FFDM manufacturer's proprietary preprocessing methods, we used the raw FFDM image as input and developed a multiresolution preprocessing scheme for image enhancement. A two-stage prescreening method that combines gradient field analysis with gray level information was developed to identify mass candidates on the processed images. The suspicious structure in each identified region was extracted by clustering-based region growing. Morphological and spatial gray-level dependence texture features were extracted for each suspicious object. Stepwise linear discriminant analysis (LDA) with simplex optimization was used to select the most useful features. Finally, rule-based and LDA classifiers were designed to differentiate masses from normal tissues. Two data sets were collected: a mass data set containing 110 cases of two-view mammograms with a total of 220 images, and a no-mass data set containing 90 cases of two-view mammograms with a total of 180 images. All cases were acquired with a GE Senographe 2000D FFDM system. The true locations of the masses were identified by an experienced radiologist. Free-response receiver operating characteristic analysis was used to evaluate the performance of the CAD system. It was found that our CAD system achieved a case-based sensitivity of 70%, 80%, and 90% at 0.72, 1.08, and 1.82 false positive (FP) marks/image on the mass data set. The FP rates on the no-mass data set were 0.85, 1.31, and 2.14 FP marks/image, respectively, at the corresponding sensitivities. This study demonstrated the usefulness of our CAD techniques for automated detection of masses on FFDM images. [DOI: 10.1118/1.1997327]

Key words: computer-aided detection, full field digital mammogram (FFDM), multiresolution image enhancement, gradient field analysis, stepwise linear discriminant analysis

I. INTRODUCTION

Breast cancer is one of the leading causes of death among American women between 40 and 55 years of age.[1] It has been reported that early diagnosis and treatment can significantly improve the chance of survival for patients with breast cancer.[2–4] Although mammography is the best available screening tool for detection of breast cancers, studies indicate that a substantial fraction of breast cancers that are visible upon retrospective analyses of the images are not detected initially.[5–8] Computer-aided diagnosis (CAD) is considered to be one of the promising approaches that may improve the sensitivity of mammography.[9,10] Computer-aided lesion detection can be used during screening to reduce oversight of suspicious lesions that warrant further work-up. Computer-aided lesion characterization can assist in the estimation of the likelihood of malignancy of lesions by using image and/or other information during the diagnostic stage. The majority of studies to date show that CAD can improve radiologists' lesion detection sensitivity,[11–16] although Gur *et al.*[17] found that CAD had no significant effect on the radiologists in their academic setting when they averaged the results from both low-volume and high-volume radiologists. Further analysis of Gur's data by Feig *et al.*[18] indicated that the 17 low-volume radiologists in Gur's study achieved similar increase in sensitivity as reported in other studies. The outcome of CAD studies therefore depends on the study design and data analysis.

A number of investigators have reported CAD algorithms for detection of masses on mammograms. Their approaches to prescreening of mass candidates were based primarily on mass characteristics including: (1) asymmetric density between left and right mammograms,[19–22] (2) texture,[23,24] (3) spiculation,[25,26] (4) gray level contrast,[27–31] and (5) gradient.[32] Some of these approaches were refined with a combination of the mass characteristics. Feature classifiers were then used to further differentiate masses from normal breast tissues.

Most mammographic CAD algorithms developed so far are based on digitized screen-film mammograms (SFMs). In the last few years, full field digital mammographic (FFDM) technology has advanced rapidly because of the potential of digital imaging to improve breast cancer detection. Several manufacturers have obtained clearance from the FDA for clinical use. It is expected that FFDM detectors will provide higher signal-to-noise ratio (SNR) and detective quantum efficiency, wider dynamic range, and higher contrast sensitivity

than digitized mammograms. The spatial resolution of digital detectors may also be different from that of digitized SFMs even when their pixel pitches are equal. Li *et al.* investigated the performance of their CAD system on mass detection that was developed for SFMs and modified for FFDMs.[33] Their preliminary results on a small data set showed that it achieved 60% sensitivity at 2.47 false positives (FPs)/image. It is expected that proper adaptation based on the imaging characteristics of FFDMs and re-training of the CAD system with FFDMs would improve the performance. Because of the higher SNR and linear response of digital detectors, there is also a strong potential that more effective feature extraction techniques can be designed to optimally extract signals from the image and improve the accuracy of CAD. Several commercial CAD systems already obtained FDA approval for use with FFDMs. The commercial CAD systems generally reported similar performance on FFDMs and SFMs. However, their study was not reported in peer-reviewed journals so that the data set and algorithm are unknown. Recently, an assessment study[34] to compare the performance of two commercial and one research CAD system for SFMs showed that their mass detection sensitivities ranged from 67% to 72% and the FP rates ranged from 1.08 to 1.68 per four-view examinations. The differences in sensitivities were not significant whereas the differences in the FP rates were significant, depending on the examinations and CAD systems used.[34]

We have developed a CAD system for the detection of masses on SFMs in our previous studies.[30,35,36] We are developing a mass detection system for mammograms acquired directly by a FFDM system. In this study, we adapted our mass detection system developed for SFMs to FFDMs by optimizing each stage and retraining. In an effort to develop a CAD system that is less dependent on the FFDM manufacturer's proprietary preprocessing methods, we used the raw FFDM as input and developed a multiresolution preprocessing scheme for image enhancement. A new technique was also designed for prescreening of mass candidates on the preprocessed images.

II. MATERIALS AND METHOD

A. Data sets

The mammograms were collected from patient files at the Department of Radiology with Institutional Review Board approval. Digital mammograms at the University of Michigan are acquired with a GE Senographe 2000D FFDM system. The GE system has a CsI phosphor/a:Si active matrix flat panel digital detector with a pixel size of 100 μm $\times$100 μm and 14 bits per pixel. In this study, we used two data sets: a mass set containing FFDMs with malignant or benign masses and a no-mass set containing FFDMs without masses. The no-mass set was obtained from microcalcification cases collected for the development of our microcalcification CAD systems. The cases were included as normal, with respect to masses, only if they were verified to be free of masses by an experienced Mammography Quality Standards Act (MQSA) radiologist. Our mass detection system aims at application to screening mammography so that the mass cases, regardless of malignant or benign, are considered positive. All cases had two mammographic views, the craniocaudal view and the mediolateral oblique view or the lateral (LM or ML) view. The mass set contained 110 cases with a total of 220 images. The no-mass set contained 90 cases with a total of 180 images. The mass data set was used to estimate the detection sensitivity and the no-mass data set was used for estimating the FP rate. There were a total of 110 biopsy-proven masses in the mass data set. Eighty-seven of the masses were benign and 23 of the masses were malignant. A MQSA radiologist identified the locations of the masses, measured the mass sizes as the longest dimension seen on the two-view mammograms, provided descriptors of the mass shapes and mass margins, and also provided an estimate of the breast density in terms of BI-RADS category. Figure 1 shows the information of our data set which includes the distributions of mass sizes, mass shapes, mass margins, and breast density.

B. Methods

Our CAD system consists of five processing steps: (1) preprocessing by using multiscale enhancement, (2) prescreening of mass candidates, (3) identification of suspicious objects, (4) feature extraction and analysis, and (5) FP reduction by classification of normal tissue structures and masses. The block diagram for the detection scheme is shown in Fig. 2. These steps are described in more detail in the following.

We randomly separated the mass data set into two independent, equal sized subsets. Each subset contained 55 cases with 110 images. Cross validation was used for training and testing the algorithms. The training included selecting the preprocessing Laplacian pyramid reconstruction weights, adjusting the filter weights for prescreening and clustering, determining thresholds for rule-based classification, and selecting morphological and texture features and classifier weights. Once the training with one subset was completed, the parameters and all thresholds were fixed for testing with the other subset. The training and test subsets were switched and the training process was repeated. The overall detection performance was evaluated by combining the performances for the two test subsets. The trained algorithms with the fixed parameters were also applied to the no-mass mammograms to estimate the FP rate in screening mammograms.

1. Preprocessing

FFDMs are generally preprocessed with proprietary methods by the manufacturer of the FFDM system before being displayed to readers. The image preprocessing method used depends on the manufacturer of the FFDM system. To develop a CAD system that is less dependent on the FFDM manufacturer″s proprietary preprocessing methods, we use the raw FFDM as input to our CAD system. We developed a multiscale preprocessing scheme for image enhancement.

Multiscale methods have been used for contrast enhancement of medical images. Since a multiscale method uses the information from a large number of frequency channels ex-

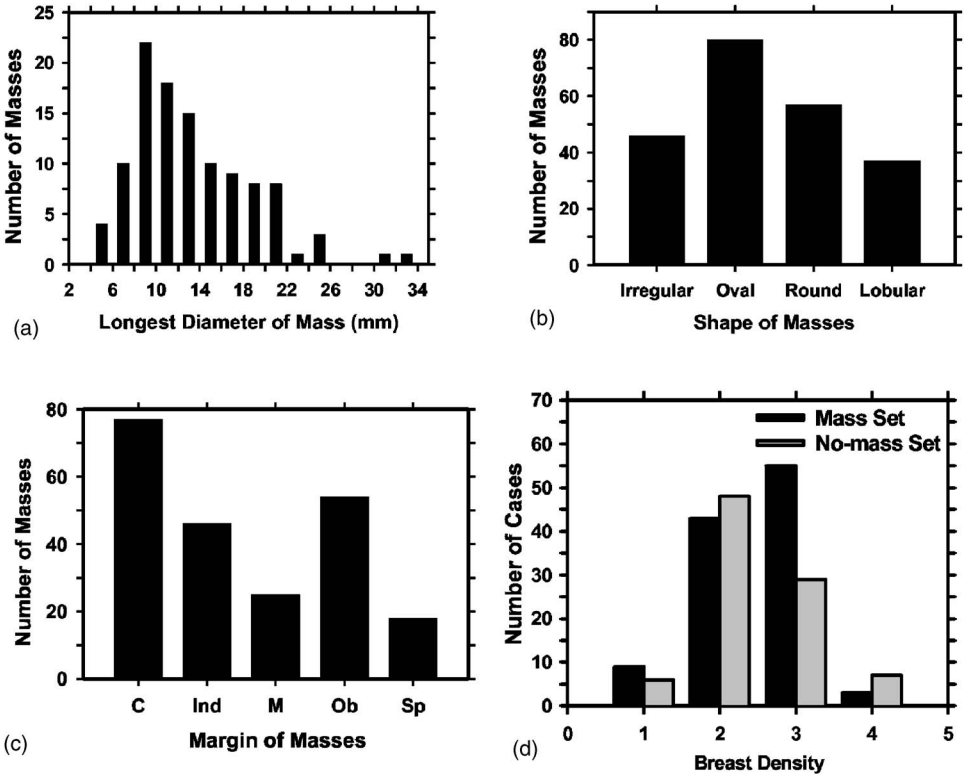


FIG. 1. The information of our mass data set: (a) distribution of mass sizes, (b) distribution of mass shapes, (c) distribution of mass margins, C: circumscribed, Ind: indistinct, M: microlobulated, Ob: obscured, Sp: spiqulated, (d) distribution of the breast density in terms of BI-RADS category estimated by a MQSA radiologist.

tracted from the image adaptively, it is more flexible and versatile than the commonly used enhancement methods, such as unsharp masking, which uses a small number of frequency channels. Two types of multiscale methods have been used as the preprocessing methods for the contrast enhancement of mammograms: the wavelet method and the Laplacian pyramid method.[37] A previous study has shown that, for the purpose of image enhancement, using a Laplacian pyramid method is advantageous compared to using the fast wavelet transformation which introduces visible artifacts.[38] In this project, therefore, we chose the Laplacian pyramid method as our preprocessing method.

A flowchart of our preprocessing method is shown in Fig. 3. In brief, the mammogram is first segmented automatically into the background and the breast region. Second, a logarithmic transform is applied to the breast image. The Laplacian pyramid method is used to decompose the breast image

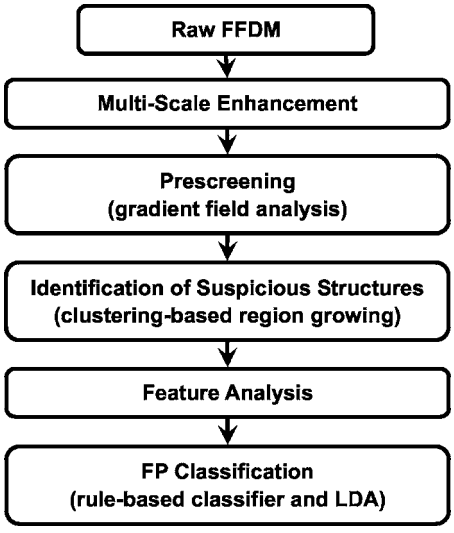


FIG. 2. Schematic diagram of our CAD system for mass detection on FFDM. The system is developed for screening mammography so that all masses, regardless of malignant or benign, are considered positive. The FP classification stage includes rule-based classification, a morphological LDA classifier, and a texture feature LDA classifier for differentiating masses from normal breast tissues.

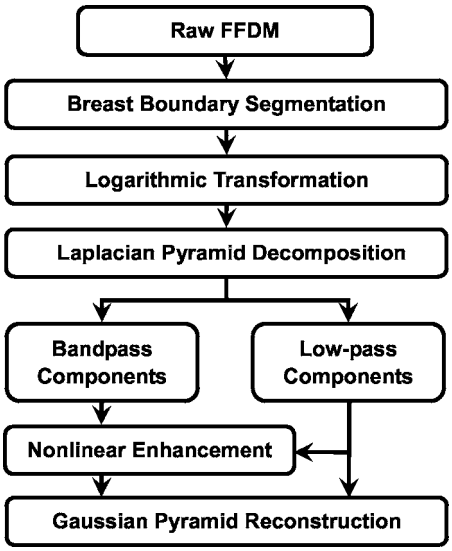


FIG. 3. Schematic diagram for the image preprocessing stage of our mass detection system, which includes breast boundary segmentation, logarithmic image transformation, and Laplacian pyramid multiscale enhancement.

into multiscales. A nonlinear weight function based on the pixel gray level from each of the low-pass components is designed to enhance the high-pass components.

Since the contrast between the breast and the background in a raw FFDM is high, a two-step algorithm was developed for the segmentation of breast region. First, Otsu's method[39] is used to calculate a threshold and binarize the original image. Second, an eight-connectivity labeling method is used to identify the connected regions below the threshold on the binary image. The region with the largest area will be considered to be the breast region.

Clinical mammograms are usually viewed in a negative mode of the raw images. In order to process an image with the same format as the clinical mammograms, we first use an inverted logarithmic function[40] to transform the raw data. A multiresolution method is then used to enhance the log-transformed image. The inverted logarithmic function for signal transfer can be expressed as

$$S_x = \ln\left(\frac{X_{\max}}{X}\right) \tag{1}$$

where X is the gray level of the raw data, $X_{\max}$ is the maximum of the 14 bit digital gray scale number (i.e., 16 383). The transformed image is then linearly scaled to 12 bit pixel values.

The Laplacian pyramid decomposition is a multiscale method that was first introduced as an image compression technique.[37] We previously evaluated the effect of Laplacian pyramid data compression on the detection of microcalcifications on digitized mammograms.[41] An illustration of a Laplacian decomposition tree is shown on the left-hand side of Fig. 4. The Laplacian pyramid is a sequence of error images $L_0, L_1, \ldots, L_n$. Each is the difference between two consecutive levels of the Gaussian pyramid $G_0, G_1, \ldots, G_n$, where G_0 is the original image. Each subsequent level of the Gaussian pyramid in the decomposition tree is generated by convolution of the image at the previous level with a 5×5 kernel, $w(m,n)$, that has weights of 0.4 at the center, 0.25 at the eight nearest neighbors of the center, and 0.05 at the 16 peripheral pixels, and then downsampled by a factor of 2, as described in Eq. (4). The decomposition of the image from level k to level $k+1$ can be expressed mathematically by

$$L_k = G_k - \text{Expand}(G_{k+1}), \tag{2}$$

where

$$\text{Expand}(G_{k+1}) = 4\sum_{m=-2}^{2}\sum_{n=-2}^{2} w(m,n)\cdot G_{k+1}\left(\frac{i-m}{2},\frac{j-n}{2}\right), \tag{3}$$

$$G_k(i,j) = \sum_{m=-2}^{2}\sum_{n=-2}^{2} w(m,n)G_{k-1}(2i+m,2j+n). \tag{4}$$

The original image can be recovered by following the Gaussian reconstruction tree shown on the right-hand side of Fig. 4 if no enhancement is applied to the Laplacian pyramid. At a given level of the Gaussian reconstruction tree, the image is expanded (convolved and upsampled), as shown in Eq. (3), and then added to the Laplacian error image of the corresponding level. Details of the decomposition and reconstruction processes can be found in the literature.[37]

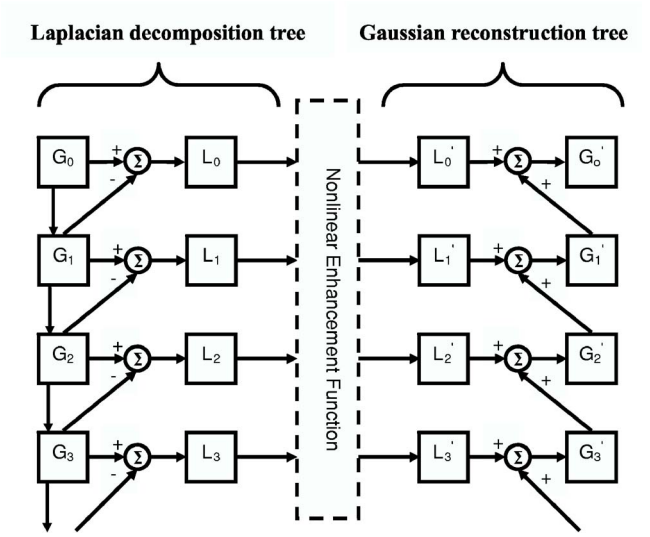


FIG. 4. Multiscale enhancement using the Laplacian pyramid decomposition method: Laplacian decomposition tree on the left-hand side and the Gaussian reconstruction tree on the right-hand side. The different levels of the Gaussian pyramid images are denoted by G_i, $(i=0,\ldots,n)$. The error images at different levels of the Laplacian pyramid are denoted by L_i, $(i=0,\ldots,n)$. The primed quantities G_i' and L_i' denoted the images at different levels after enhancement. $\sum$ denotes the summation operation. The image is downsampled by a factor of 2 when it goes down every level of the decomposition tree, and upsampled by a factor 2 when it moves up every level of the reconstruction tree.

We enhance the reconstructed image to facilitate mass detection. The image at each level of the Laplacian pyramid that corresponds to a bandpass image is mapped by a nonlinear function. In this study, we use a nonlinear function that incorporates the information from each bandpass image. A Gaussian pyramid expansion is then used to reconstruct the image from the low pass components and the enhanced bandpass components, as shown in Fig. 4. The reconstruction scheme is defined by

$$r(k) = \alpha\cdot\text{Expand}(G_{k+1}) + \beta\cdot(\text{Expand}(G_{k+1}))^p\cdot L_k, \tag{5}$$

where α, β, and p are constant values in the range of 0.2–2.0 experimentally chosen for each frequency level.

Figures 5(a) and 5(b) show an example of a GE raw image and its processed image provided by the GE FFDM system. The histograms of the raw image and the processed image are shown next to the corresponding images. An example of the processed image using our multiresolution enhancement method and the corresponding histogram are shown in Fig. 5(c).

2. Prescreening and segmentation of suspicious objects

In our previous CAD system developed for digitized SFMs, an adaptive density-weighted contrast enhancement (DWCE) filter[35] was developed for prescreening. Although

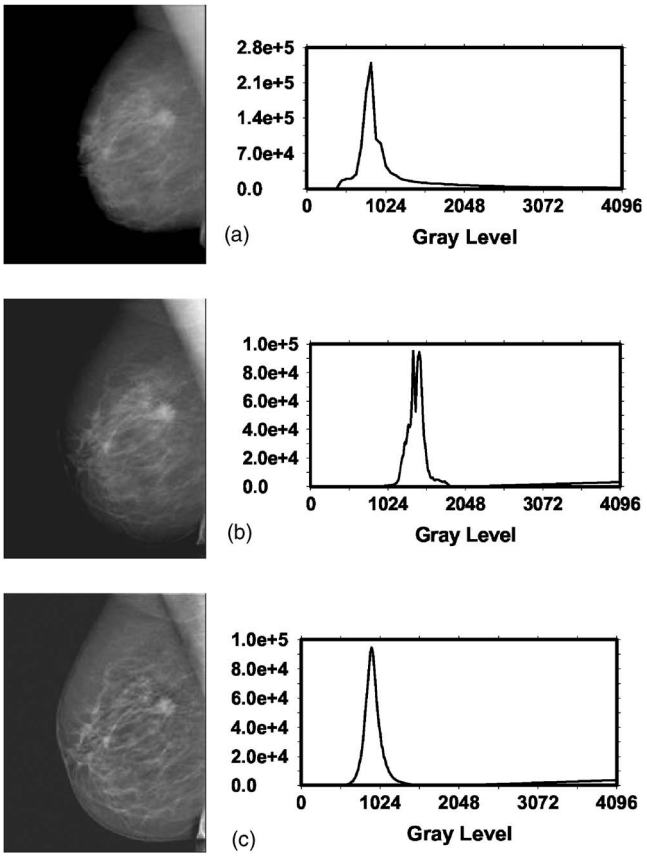


FIG. 5. An example of (a) GE raw image, (b) GE processed image, and (c) our processed image by using the Laplacian pyramid multiscale method. The gray level histogram of each image is also shown. The GE raw image has 14 bit gray levels but the histogram only plotted the lower 12 bits because very few pixels had gray levels higher than 4095.

the DWCE filter using the gray level information can identify the suspicious locations of masses on mammograms with high sensitivity, the prescreening objects often include a large number of enhanced normal breast structures.

In this study, we investigated the use of a new method that combines gradient field information and gray level information to detect mass candidates on FFDMs. Gradient field information is commonly used in computer vision or other fields to extract objects or intensity field distributions. Kobatake *et al.*[42] designed a filter, referred to as an iris filter, to calculate the convergence of gradient index around each pixel on SFMs which provided shape information for detection of masses. An extension of the iris filter, referred to as an adaptive ring filter, was developed by Wei *et al.*[43] for detection of lung nodules on chest x-ray images. In this study, we have developed a two-stage gradient field analysis method which uses not only the shape information of masses on mammograms but also incorporates the gray level information of the local object segmented by a region growing technique in the second stage to refine the gradient field analysis.

To reduce noise in the gradient calculation, the image is smoothed with a 4×4 box filter and subsampled to 400 μm $\times$ 400 μm. The gradient field analysis is applied to the smoothed image. At each pixel $c(i)$ within the breast, concentric annular regions centered at $c(i)$ with an average radius, $R(k)$, of k pixels from $c(i)$ and a radial width of 4 pixels are defined within a circular region of about 12 mm in radius. The gradient vector at each pixel $p(j)$ within an annular region is computed and the gradient direction is obtained by projecting the gradient vector to the radial direction vector from $c(i)$ to $p(j)$. The average gradient direction over an annular region at the average radius $R(k)$ is calculated as the mean of the gradient directions over pixels on three adjacent annular regions $R(k-1)$, $R(k)$, and $R(k+1)$. Finally, the gradient field convergence at $c(i)$ was determined as the maximum of the average gradient directions among all annular regions. A region of interest (ROI) of 256×256 pixels in the 100 μm $\times$ 100 μm images is identified with its center placed at each location of high gradient convergence. The object in each ROI is segmented by a region growing method[44] in which the location of high gradient convergence is used as the starting point. After region growing, all connected pixels constituting the object are labeled. Finally, the gradient convergence at the center location of the ROI is recalculated within the segmented object. Objects whose new gradient convergence is lower than 80% of the original value are rejected.

After prescreening, the suspicious objects are identified by using a two-stage segmentation method. First, the background-corrected ROI was weighted by a Gaussian function with $\sigma=256$ pixels. Then, a k-means clustering using the pixel values in a background-corrected image and a Sobel filtered image as features is used to find the object. Figures 6(a) and 6(b) show the initial detection locations and the grown objects, respectively, obtained by prescreening the mammogram shown in Fig. 5(c).

3. Feature extraction and FP reduction

FP classification in our mass detection system is accomplished by a three-stage classification scheme.[36,44] For each suspicious object, eleven morphological features are extracted. Rule-based classification and a linear discriminant analysis (LDA) classifier using all 11 morphological features as input predictor variables are trained to remove the detected structures that are substantially different from breast masses. The training data set alone was used for training the classification rules and the weights of the LDA classifier. After morphological classification, global and local multiresolution texture analyses[45] are performed in each remaining ROI by using the spatial gray level dependence (SGLD) matrix. Briefly, the wavelet transform is employed to decompose an ROI into three levels for global texture analysis. Thirteen types of texture features[44,46] are extracted from each ROI. Each feature is calculated at 14 pixel distances and 2 angular directions. A total of 364 features (13 texture measures $\times$ 14 distances $\times$ 2 directions) is extracted from global texture analysis. Local texture features are extracted from the local region containing the detected object (object region) and the peripheral regions within each ROI. A total of 208 features (104 features from the object region and 104

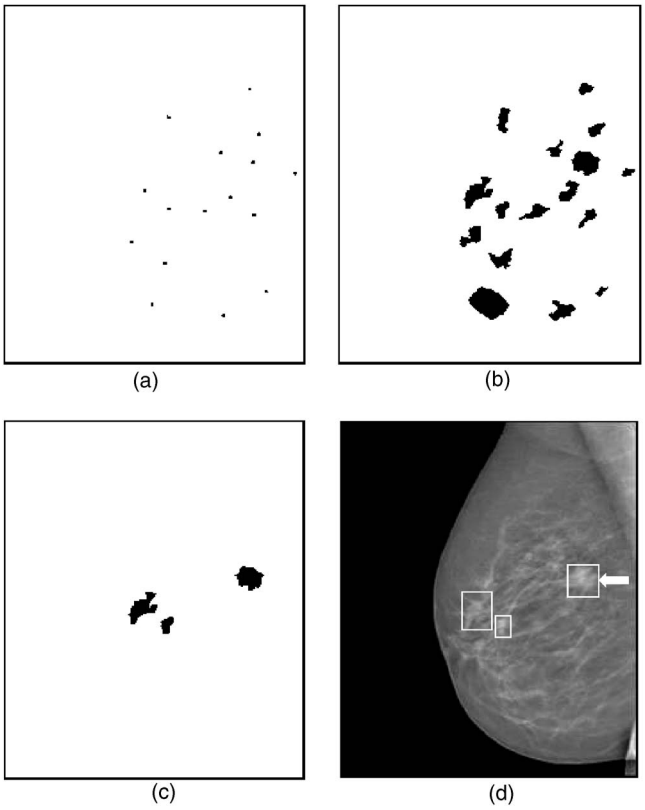


FIG. 6. An example demonstrating the processing steps with our CAD system: (a) object locations identified in prescreening, (b) identified suspicious objects, (c) detected objects after FP reduction, and (d) image superimposed with ROIs identifying the detected objects. The true mass is indicated by an arrow.

features from the peripheral regions) are extracted. The third-stage FP reduction using the texture features is described next.

4. Texture classification of masses and normal tissue

In order to obtain the best texture feature subset and reduce the dimensionality of the feature space to design an effective classifier, feature selection with stepwise LDA was applied. At each step one feature was entered or removed from the feature pool by analyzing its effect on the selection criterion, which was chosen to be the Wilks' lambda in this study. The optimization procedure used a threshold F_{in} for feature entry, a threshold F_{out} for feature removal, and a tolerance threshold T for excluding features that had high correlation with the features already in the selected pool. Since the appropriate values of F_{in}, F_{out}, and T were unknown, we examined a range of F_{in}, F_{out}, and T values using an automated simplex optimization method. For a given combination of F_{in}, F_{out}, and T values, the algorithm used a leave-one-case-out resampling method within the training subset to select features and estimate the weights for the LDA classifier. To evaluate the classifier performance, the test discriminant scores from the left-out cases were analyzed using receiver operating characteristic (ROC) methodology.[47] The discriminant scores of the mass and normal tissue were used as the decision variable in the LABROC program, which fits a binormal ROC curve based on maximum likelihood estimation. The accuracy for classification of mass and normal tissue was evaluated as the area under the ROC curve, A_z. The test A_z for the left-out cases in the leave-one-out resampling within the training subset was used as a figure of merit to guide the simplex algorithm to search for the best set of F_{in}, F_{out}, and T values within the parameter space. In this approach, feature selection was performed without the left-out case so that the test performance would be less optimistically biased.[48] However, the selected feature set in each leave-one-case-out cycle could be slightly different because every cycle had one training case different from the other cycles. In order to obtain a single trained classifier to apply to the test subset, a final stepwise feature selection was performed with the entire training subset and a set of F_{in}, F_{out}, and T thresholds chosen from the output of simplex training process. This set of F_{in}, F_{out}, and T thresholds was chosen based not only on the test A_z values, which were generated when the simplex procedure was searching through the parameter space, but also on the average number of features selected. The appropriate thresholds were chosen as a balance between keeping the number of selected features small and a relatively high classification accuracy by LDA. The chosen thresholds were then applied to the entire training subset to obtain the final set of features using stepwise feature selection and estimate the weights of the LDA. The LDA classifier with the selected feature set was then fixed and applied to the test subset. The test subset was independent of the training subset as described in Sec. II B 2 and was not used in the above-described leave-one-case-out classifier training process.

5. Evaluation methods

The detected individual objects were compared with the "truth" ROI marked by an experienced radiologist. A detected object was scored as true positive (TP) if the overlap between the bounding box of the detected object and the truth ROI was over 25%. Otherwise, it would be scored as FP. The 25% threshold was selected as described in our previous study.[36] The detection performance of the CAD system was assessed by free response ROC (FROC) analysis. FROC curves were presented on a per-mammogram and a per-case basis. For mammogram-based FROC analysis, the mass on each mammogram was considered an independent true object; the sensitivity was thus calculated relative to 220 masses. For case-based FROC analysis, the same mass imaged on the two-view mammograms was considered to be one true object and detection of either or both masses on the two views was considered to be a TP detection; the sensitivity was thus calculated relative to 110 masses. Figure 6(c) shows an example of the final detected objects and Fig. 6(d) shows the locations of these objects superimposed on the mammogram.

To evaluate the effect of the preprocessing methods on mass detection, we also trained a CAD system using the GE

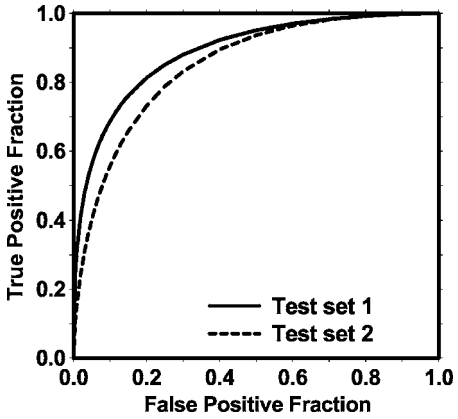


FIG. 7. The test ROC curves from the two independent mass subsets. The LDA classifiers using text features achieved an A_z value of 0.89±0.02 for test subset 1 and 0.85±0.02 for test subset 2 in the classification of mass and normal breast tissues.

processed image as input. This CAD system used the same methods as those described earlier for the raw images except that the Laplacian pyramid preprocessing step was not applied to the GE processed image, and that the prescreening and feature classifiers were retrained specifically for the GE processed images to obtain the best performance. The training and test subsets contained the same corresponding cases as for the raw image subsets. The training and testing were performed using the above-described cross validation method. The performance of the CAD system using the GE processed images was quantified by the average test FROC curve and compared with that using the raw images.

III. RESULTS

With raw images as input and Laplacian pyramid enhancement, our CAD system using the two-stage gradient field analysis detected 92.7% (204/220) of the masses with an average of 18.9 (4152/220) objects/image at the prescreening stage, compared with an average of 23.8 objects/image at the same sensitivity by using gradient field information alone. After FP reduction using the rule-based and linear classifier based on morphological features, there were a total of 3412 mass candidates (15.5 objects/image) at a sensitivity of 90.5% (199/220).

The texture-based LDA classifier for FP reduction was designed with stepwise feature selection and simplex optimization. The most effective subset of features from the available feature pool was selected for each of the training subsets during the training procedure. Twenty (11 global and 9 local) and 19 (12 global and 7 local) texture features were selected from the two independent training subsets, respectively. The test ROC curves are shown in Fig. 7. The training A_z values of the LDA classifier on the two training subsets were 0.87±0.02 and 0.88±0.01, respectively. The classifiers achieved A_z values of 0.89±0.02 and 0.85±0.02 on the independent test subsets, respectively. Figure 8 shows the FROC curves for the two test subsets after FP reduction with the corresponding trained LDA classifiers. An average FROC curve was derived from these two FROC curves by averaging the FP/images at the corresponding sensitivities. This average test FROC curve is plotted in Fig. 9 for comparison with the other FROC curves, described next.

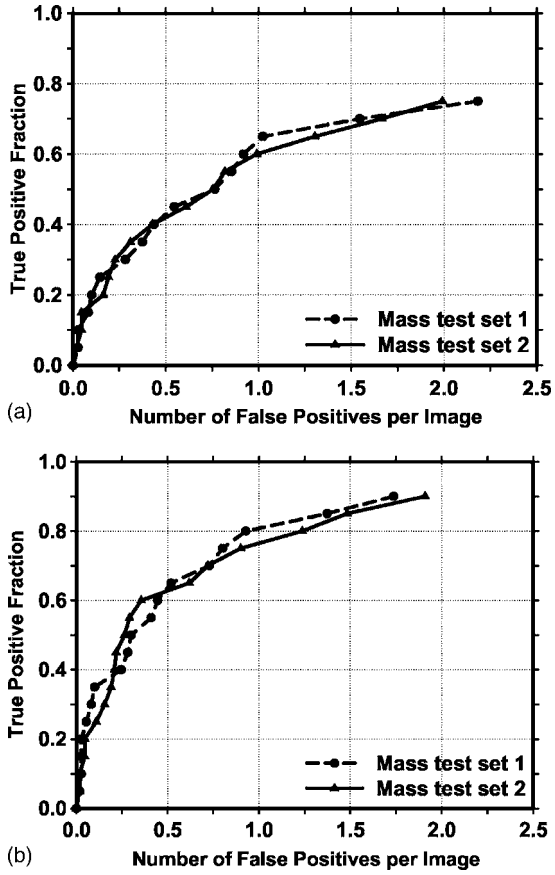


FIG. 8. The test FROC curves from the two independent mass subsets for the CAD system using the raw images as input and processed with the Laplacian pyramid method. The FP rate was estimated from the mammograms with masses. (a) Image-based FROC curves, (b) case-based FROC curves.

In addition to using the mass data set containing 110 cases for the cross validation training and testing, we used a no-mass data set containing 90 cases with 180 images to evaluate the FP detection rate in normal cases. Since two sets of trained parameters were acquired as a result of the cross validation training, we applied the two trained CAD systems separately to the no-mass data set for FP detection. The number of FP marks produced by the algorithm was determined by counting the detected objects on these normal cases only. The mass detection sensitivity was determined by counting only the abnormal objects on each of the test mass subsets. The combination of the sensitivity from each of the test mass subsets and the FP rate from the normal data set at the corresponding detection thresholds resulted in a test FROC curve. The two test FROC curves were then averaged, as described earlier, to obtain an overall FROC curve quantifying the test performance of the CAD system. Figures 9(a) and 9(b) show the comparison of the average FROC curves with the FP rates estimated from the two data sets. The test FROC curve with the FP rate estimated from the no-mass data set showed a case-based detection sensitivity of 70%,

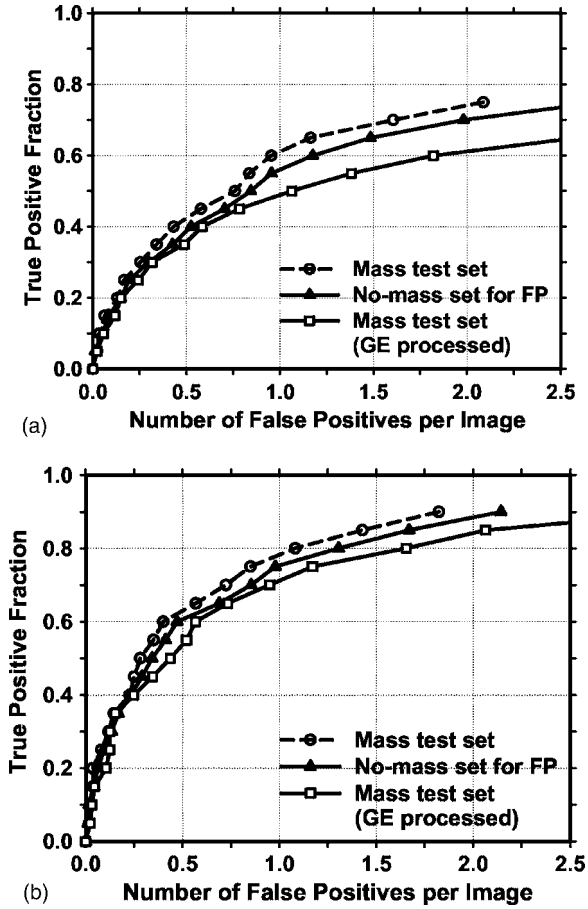


FIG. 9. Comparison of the average test FROC curves obtained from: (1) the CAD system using raw images as input, with the FP rate estimated from the mammograms with masses, (2) the CAD system using raw images as input, with the FP rate estimated from the normal mammograms without masses, and (3) the CAD system using GE processed images as input, with the FP rate estimated from the GE processed mammograms with masses. (a) Image-based FROC curves, (b) case-based FROC curves.

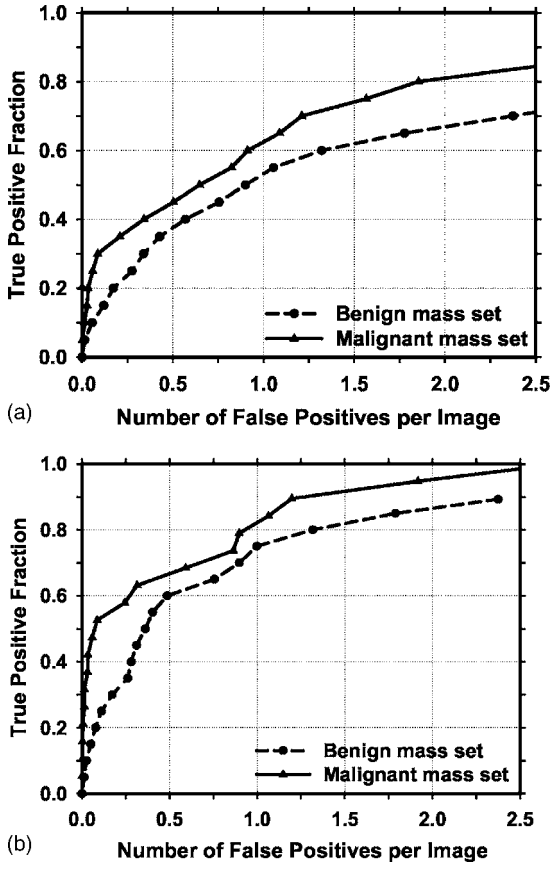


FIG. 10. Comparison of the average test FROC curves for the malignant and benign mass sets. The CAD system using raw images as input was used and the FP rate was estimated from the mammograms without masses. (a) Image-based FROC curves, (b) case-based FROC curves.

80%, and 90% at 0.85, 1.31, and 2.14 FP marks/image, which are slightly higher than the FP rates of 0.7, 1.1, and 1.8 marks/image, respectively, estimated from the mass data set. Since our mass detection algorithm limits the maximum number of output marks to be 3 at the final stage, the FP marker rates will be slightly higher if the detection is performed in no-mass images. However, many images do not reach the maximum of 3 marks so that the difference in the FP marker rate between the mass and no-mass set is less than one. We also analyzed the detection accuracy of the system for malignant and benign masses separately. Figures 10(a) and 10(b) show the average FROC curves for detection of malignant and benign masses.

The average test FROC curves of the CAD system using the GE processed images as input were compared to those of the CAD system using raw images as input and Laplacian pyramid multiscale preprocessing as shown in Fig. 9. The FROC curves were plotted as the detection sensitivity as a function of the number of FP marks per image on the mass data set. The CAD system using the GE processed images as input achieved a case-based sensitivity of 70%, 80%, and 90% at 0.9, 1.6, and 3.1 FP marks/image, respectively, compared with 0.7, 1.1, and 1.8 FP marks/image on the CAD system using raw images as input.

IV. DISCUSSION

Several FFDM systems have been approved for clinical applications. It is important to develop a CAD system that can easily be adapted to images acquired by FFDM systems from different manufacturers. In this study, we are developing a CAD system that uses the raw FFDMs as the input. Since digital detectors generally have a linear response to x-ray exposure, the raw pixel values are a linear function of the absorbed x-ray energy in the detector. The signal range between different digital detectors can therefore be normalized linearly with respect to each other. Although the spatial resolution and noise properties of the images from different detectors are still different, the use of raw images already reduces one of the major differences between mammograms from different FFDM systems. For preprocessing of the raw images, we developed a multiresolution enhancement method. An example of a typical mammogram processed by the GE method and our method is compared in Fig. 5. As seen from this example, the enhancement of mammographic structures was stronger for our processed image than for the

TABLE I. Estimation of the statistical significance in the difference between the FROC performance of the CAD system using the FFDM raw images as input and processed with our Laplacian pyramid method and that of the CAD system using GE processed images as input. The FROC curves with the FP rates obtained from the no-mass data set (Fig. 9) were compared.

	A_1 (AFROC)			FOM (JAFROC)		
	Test subset 1	Test subset 2	p values	Test subset 1	Test subset 2	p values
Raw+LP processed	0.44	0.39	0.012	0.46	0.41	0.006
GE processed	0.37	0.31	0.0009	0.39	0.34	0.012

GE processed image. From a comparison of their histograms, it was found that the two histograms are very similar except for the average gray level.

For the evaluation of the effect of the preprocessing methods on computerized mass detection, we observed that our Laplacian pyramid preprocessing method provided higher detection accuracy than the GE processing method. As shown in Fig. 5, the Laplacian pyramid preprocessing method applies a stronger edge enhancement to the image than the GE method. Our preprocessing method aims at enhancing the image structures for computer vision whereas the GE processing method was designed to enhance the image for human visual interpretation. The stronger enhancement used for preprocessing the raw images appeared to improve the accuracy of the computer in detecting the masses.

Currently, there is no established statistical analysis method for testing the significance of the difference between two FROC curves generated by a CAD system. Chakraborty *et al.* proposed using an alternative free-response ROC (AFROC) method[49] to transform the FROC data to AFROC data, to which the curve fitting software and statistical significance tests for ROC analysis can then be applied and demonstrated its application to human observer performance rating data. In the AFROC method, false-positive images (FPIs) instead of FPs per image are counted. The confidence rating of a FPI is determined by the highest confidence FP decision on the image regardless of how many lower confidence FP decisions are made on the same image. We applied the AFROC method to evaluate the differences in pairs of our FROC curves that used the no-mass set for estimation of the FP rates. The ROCKIT software developed by Metz *et al.*[47] was used to analyze the AFROC data. The comparison of A_1 and p values is summarized in Table I. The area under the fitted AFROC curve (A_1) was 0.44 and 0.39, respectively, on mass test subsets 1 and 2 for the CAD system using raw images as input and processed with our Laplacian pyramid method, and 0.37 and 0.31, respectively, on the same subsets for the CAD system using GE processed images as input. The difference between the fitted AFROC curve for our processed images and that for the GE processed images was statistically significant ($p<0.05$) for both test subsets. However, all four fitted AFROC curves deviated systematically from the AFROC data (see two examples plotted in Fig. 11 for the test subset 1). It is uncertain whether the AFROC method is applicable to our FROC data and thus whether the statistical significance testing is valid.

More recently, Chakraborty *et al.*[50] proposed a JAFROC method and provided software to estimate the statistical significance of the difference between two FROC curves. We also applied the JAFROC analysis to the two pairs of FROC curves. The figure-of-merit (FOM) from the output of the JAFROC software was 0.46 and 0.41, respectively, on mass test subsets 1 and 2 for the CAD system using raw images as input and processed with our Laplacian pyramid method, and 0.39 and 0.34, respectively, on the same subsets for the CAD system using GE processed images as input. The difference between the FOM for our processed images and that for the GE processed images was again statistically significant ($p<0.05$). The FOM values were about 0.02 higher than the corresponding A_1 values. The JAFROC software did not provide a fitted curve or a goodness-of-fit indicator in the output so that it is not known whether this model fits our FROC data better than the AFRPC method. Although both methods indicate that the improvement in the FROC performance using our Laplacian pyramid processed images is statistically

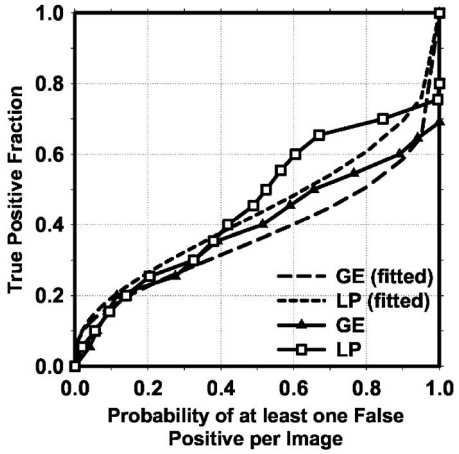


FIG. 11. Comparison of alternative free-response receiver operating characteristic (AFROC) curves. The raw curves were transformed from the FROC curves of mass detection on test subset 1 using either the raw images as input and processed with the Laplacian pyramid method (LP) or the GE processed images as input. The FP rate was estimated from the mammograms without masses. The fitted AFROC curves were obtained by applying the ROCKIT program to the transformed AFROC data.

significant, further investigations are needed to study whether these models are valid for analyzing the FROC performance of CAD systems.

The prescreening technique is an important task in a CAD system. A number of researchers have developed methods for detection of suspicious masses on SFMs and CRs. The previous methods produced between 10 to 30 FPs/image for a mass detection sensitivity of approximately 90%. However, it is difficult to compare the effectiveness of the different methods because of the differences in the image recording systems and in the data sets. In this study, we developed a new method that combines gradient field information, which was originally developed for the detection of lung nodules on chest x-ray images,[43] and gray level information[44] for prescreening mass candidates on the FFDMs. The new method produced 18.9 objects/image at 93% sensitivity in the prescreening step, compared with an average of 23.8 objects/image at the same sensitivity by using gradient field information alone.

The texture features in this study were extracted by using the SGLD matrix. A total of 572 features were included in our initial feature pool. These features were also used by our CAD system previously developed for SFMs. An average number of 19.5 features were selected by using a stepwise feature selection method. The A_z values for the LDA classifiers were 0.87 ± 0.02 and 0.88 ± 0.01 on the two training subsets, and 0.89 ± 0.02 and 0.85 ± 0.02 on the test subsets, respectively. The slightly higher test A_z from the first test subset than the A_z from its training subset may indicate that some relatively easy cases were assigned, by chance, to that test set during random partitioning. We also investigated if other features could improve the performance of our CAD system. The different feature spaces that we examined included features extracted from principal component analysis applied to the ROI image, run length statistics texture features extracted from the ROI images, and combination of one or both of these feature spaces with the SGLD feature space. However, the test results showed that a LDA classifier designed in the SGLD feature space alone provided the best performance. Although this was found to be true for both our CAD mass detection system for SFMs developed previously and the current system for FFDMs, it is still difficult to conclude that the SGLD features are the best feature set for classification between breast masses and normal tissues. One major concern of the SGLD feature space is that the dependence of the feature values on the pixel pair distance and angular direction leads to a feature pool with a large number of features. Some features in such a large feature space may provide good performance in classification of masses and normal structures by chance. We attempted to alleviate this problem by using an independent test set to evaluate the classifier performance. However, since we chose the overall system parameters with the knowledge of the performance for the test sets, the evaluation would still amount to validation rather than true testing. We have verified that our CAD system for SFMs can achieve reasonable performance in a true independent data set[36] and a prospective pilot clinical trial.[16] The performance of the current CAD system for FFDMs will have to be evaluated similarly when independent data sets become available.

The detection performance of a CAD system for malignant masses is more important than its performance for all masses. Figures 10(a) and 10(b) indicate that the sensitivity of the system is higher for malignant masses than for benign masses. This is consistent with our observation in previous studies of our CAD system for digitized SFMs.[36] However, since our current data set contained only 23 malignant cases, there will be large statistical uncertainty in the evaluation of sensitivity in this subset. A larger data set is being collected for comparing the detection performances of the CAD system between malignant and benign masses and also for the purpose of classifying malignant and benign masses. Furthermore, CAD algorithms developed for SFMs have been proven to be useful as a second opinion to assist radiologists in mammographic interpretation. Because of the higher SNR and linear response of digital detectors, there is also a potential that FFDMs can improve the sensitivity of breast cancer detection, especially in dense breasts. Several studies have been or are being conducted to compare FFDM with SFM in screening cohorts. It is also important to compare the performance of CAD systems between FFDMs and SFMs. A study is under way to compare the performance of the two systems on pairs of FFDM and SFM obtained from the same patients.[51]

V. CONCLUSION

Several FFDM systems have been approved for clinical applications. It is important to develop CAD systems for breast cancer detection in FFDM. In this work, we developed a CAD system that uses the raw FFDMs as the input. A multiresolution Laplacian pyramid enhancement method was devised to preprocess the raw FFDMs. A new prescreening method that combined gradient field analysis with gray level information was developed to identify mass candidates. Rule-based and LDA classifiers in a feature space which consisted of morphological features and SGLD texture features were designed to differentiate masses from normal tissues. It was found that our CAD system achieved a case-based sensitivity of 70%, 80%, and 90% with an estimate of 0.85, 1.31, and 2.14 FP marks/image, respectively, on normal cases. The results indicate that our mass detection CAD scheme can be useful for detecting masses on FFDMs. Studies are under way to further optimize the processing parameters, the feature extraction, and the classifiers for FP reduction. Comparison of mass detection performance of our CAD system for FFDMs and that for SFMs is also in progress.

ACKNOWLEDGMENTS

This work is supported by USPHS Grant No. CA95153, U. S. Army Medical Research and Material Command Grant Nos. DAMD 17-02-1-0214 and W81XWH-04-1-0475. The content of this paper does not necessarily reflect the position of the government and no official endorsement of any equipment and product of any companies mentioned should be

inferred. The authors are grateful to Charles E. Metz, Ph.D., for the LABROC and ROCKIT programs.

[a]Electronic mail: jvwei@umich.edu

[1]American Cancer Society, www.cancer.org 2004, "Statistics for 2004."
[2]C. R. Smart, R. E. Hendrick, J. H. Rutledge, and R. A. Smith, "Benefit of mammography screening in women ages 40 to 49 years: Current evidence from randomized controlled trials," Cancer (N.Y.) **75**, 1619–1626 (1995).
[3]S. A. Feig, C. J. D'Orsi, R. E. Hendrick, V. P. Jackson, D. B. Kopans, B. Monsees, E. A. Sickles, C. B. Stelling, M. Zinninger, and P. Wilcox-Buchalla, "American College of Radiology guidelines for breast cancer screening," Am. J. Roentgenol **171**, 29–33 (1998).
[4]B. Cady and J. S. Michaelson, "The life-sparing potential of mammographic screening," Cancer (N.Y.) **91**, 1699–1703 (2001).
[5]B. J. Hillman, L. L. Fajardo, T. B. Hunter, B. Mockbee, C. E. Cook, R. M. Hagaman, J. C. Bjelland, C. S. Frey, and C. J. Harris, "Mammogram interpretation by physician assistants," Am. J. Roentgenol. **149**, 907–911 (1987).
[6]C. A. Beam, P. M. Layde, and D. C. Sullivan, "Variability in the interpretation of screening mammograms by US radiologists—Findings from a national sample," Arch. Intern Med. **156**, 209–213 (1996).
[7]R. L. Birdwell, D. M. Ikeda, K. F. O'Shaughnessy, and E. A. Sickles, "Mammographic characteristics of 115 missed cancers later detected with screening mammography and the potential utility of computer-aided detection," Radiology **219**, 192–202 (2001).
[8]J. G. Elmore, C. Y. Nakano, T. D. Koepsell, L. M. Desnick, C. J. D'Orsi, and D. F. Ransohoff, "International variation in screening mammography interpretations in community-based programs," J. Natl. Cancer Inst. **95**, 1384–1393 (2003).
[9]F. Shtern, C. Stelling, B. Goldberg, and R. Hawkins, "Novel technologies in breast imaging: National Cancer Institute perspective," Orlando, FL.
[10]C. J. Vyborny, "Can computers help radiologists read mammograms?," Radiology **191**, 315–317 (1994).
[11]H. P. Chan, K. Doi, C. J. Vyborny, R. A. Schmidt, C. E. Metz, K. L. Lam, T. Ogura, Y. Wu, and H. MacMahon, "Improvement in radiologists' detection of clustered microcalcifications on mammograms. The potential of computer-aided diagnosis," Invest. Radiol. **25**, 1102–1110 (1990).
[12]L. J. Warren Burhenne, S. A. Wood, C. J. D'Orsi, S. A. Feig, D. B. Kopans, K. F. O'Shaughnessy, E. A. Sickles, L. Tabar, C. J. Vyborny, and R. A. Castellino, "Potential contribution of computer-aided detection to the sensitivity of screening mammography," Radiology **215**, 554–562 (2000).
[13]T. W. Freer and M. J. Ulissey, "Screening mammography with computer-aided detection: Prospective study of 12,860 patients in a community breast center," Radiology **220**, 781–786 (2001).
[14]R. F. Brem, J. K. Baum, M. Lechner, S. Kaplan, S. Souders, L. G. Naul, and J. Hoffmeister, "Improvement in sensitivity of screening mammography with computer-aided detection: A multi-institutional trial," Am. J. Roentgenology **181**, 687–693 (2003).
[15]S. V. Destounis, P. DiNitto, W. Logan-Young, E. Bonaccio, M. L. Zuley, and K. M. Willison, "Can computer-aided detection with double reading of screening mammograms help decrease the false-negative rate? Initial experience," Radiology **232**, 578–584 (2004).
[16]M. A. Helvie, L. M. Hadjiiski, E. Makariou, H. P. Chan, N. Petrick, B. Sahiner, S. C. B. Lo, M. Freedman, D. Adler, J. Bailey *et al.*, "Sensitivity of noncommercial computer-aided detection system for mammographic breast cancer detection—A pilot clinical trial," Radiology **231**, 208–214 (2004).
[17]D. Gur, J. H. Sumkin, and H. E. Rockette, "Response to Re: Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system," J. Natl. Cancer Inst. **96**, 1261 (2004).
[18]S. A. Feig, E. A. Sickles, W. P. Evans, and M. N. Linver, "Re. Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system," J. Natl. Cancer Inst. **96**, 1260–1261 (2004).
[19]F. Winsberg, M. Elkin, J. Macy, V. Bordaz, and W. Weymouth, "Detection of radiographic abnormalities in mammograms by means of optical scanning and computer analysis," Radiology **89**, 211–115 (1967).
[20]J. L. Semmlow, A. Shadagopappan, L. V. Ackerman, W. Hand, and F. S. Alcorn, "A fully automated system for screening mammograms," Comput. Biomed. Res. **13**, 350–362 (1980).
[21]T. K. Lau and W. F. Bischof, "Automated detection of breast tumors using the asymmetry approach," Comput. Biomed. Res. **24**, 273–295 (1991).
[22]F. F. Yin, M. L. Giger, K. Doi, C. E. Metz, C. J. Vyborny, and R. A. Schmidt, "Computerized detection of masses in digital mammograms: Analysis of bilateral subtraction images," Med. Phys. **18**, 955–963 (1991).
[23]C. Kimme, B. J. O'Laughlin, and J. Sklansky, *Automatic Detection of Suspicious Abnormalities in Breast Radiographs*, Data Structures, Computer Graphics and Pattern Recognition (Academic, New York, 1975).
[24]W. P. Kegelmeyer, J. M. Pruneda, P. D. Bourland, A. Hillis, M. W. Riggs, and M. L. Nipper, "Computer-aided mammographic screening for spiculated lesions," Radiology **191**, 331–337 (1994).
[25]S. L. Ng and W. F. Bischof, "Automated detection and classification of breast tumors," Comput. Biomed. Res. **25**, 218–237 (1992).
[26]N. Karssemeijer and G. te Brake, "Detection of stellate distortions in mammograms," IEEE Trans. Med. Imaging **15**, 611–619 (1996).
[27]S. M. Lai, X. Li, and W. F. Bischof, "On techniques for detecting circumscribed masses in mammograms," IEEE Trans. Med. Imaging **8**, 377–386 (1989).
[28]D. Brzakovic, X. M. Luo, and P. Brzakovic, "An approach to automated detection of tumors in mammograms," IEEE Trans. Med. Imaging **9**, 233–241 (1990).
[29]A. F. Laine, S. Schuler, J. Fan, and W. Huda, "Mammographic feature enhancement by multiscale analysis," IEEE Trans. Med. Imaging **13**, 725–740 (1994).
[30]N. Petrick, H. P. Chan, D. Wei, B. Sahiner, M. A. Helvie, and D. D. Adler, "Automated detection of breast masses on mammograms using adaptive contrast enhancement and texture classification," Med. Phys. **23**, 1685–1696 (1996).
[31]B. Zheng, Y. H. Chang, and D. Gur, "Computerized detection of masses in digitized mammograms using single-image segmentation and a multilayer topographic feature analysis," Acad. Radiol. **2**, 959–966 (1995).
[32]H. Kobatake, M. Murakami, H. Takeo, and S. Nawano, "Computer detection of malignant tumors on digital mammograms," IEEE Trans. Med. Imaging **18**, 369–378 (1999).
[33]L. Li, R. A. Clark, and J. A. Thomas, "Computer-aided diagnosis of masses with full-field digital mammography," Acad. Radiol. **9**, 4–12 (2002).
[34]D. Gur, J. S. Stalder, L. A. Hardesty, B. Zheng, J. H. Sumkin, D. M. Chough, B. E. Shindel, and H. E. Rockette, "Computer-aided detection performance in mammographic examination of masses: Assessment," Radiology **233**, 418–423 (2004).
[35]N. Petrick, H. P. Chan, B. Sahiner, and D. Wei, "An adaptive density-weighted contrast enhancement filter for mammographic breast mass detection," IEEE Trans. Med. Imaging **15**, 59–67 (1996).
[36]N. Petrick, H. P. Chan, B. Sahiner, M. A. Helvie, S. Paquerault, and L. M. Hadjiiski, "Breast cancer detection: Evalution of a mass detection algorithm for computer-aided diagnosis: Experience in 263 patients," Radiology **224**, 217–224 (2002).
[37]P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," IEEE Trans. Commun. **COM-31**, 337–345 (1983).
[38]S. Dippel, M. Stahl, R. Wiemker, and T. Blaffert, "Multiscale constrast enhancement for radiographies: laplacian pyramid versus fast wavelet transform," IEEE Trans. Med. Imaging **21**, 343–353 (2002).
[39]N. Otsu, "A threshold selection method from gray-level histograms," IEEE Trans. Syst. Man Cybern. **9**, 62–66 (1979).
[40]A. Burgess, "On the noise variance of a digital mammography system," Med. Phys. **31**, 1987–1995 (2004).
[41]H. P. Chan, S. C. B. Lo, L. T. Niklason, D. M. Ikeda, and K. L. Lam, "Image compression in digital mammography: Effects on computerized detection of subtle microcalcifications," Med. Phys. **23**, 1325–1336 (1996).
[42]H. Kobatake and S. Hashimoto, "Convergence index filter for vector fields," IEEE Trans. Image Process. **8**, 1029–1038 (1999).
[43]J. Wei, Y. Hagihara, and H. Kobatake, "Detection of rounded opacities on chest radiographs using convergence index filter," ICIAP 99, Venice, 27–29 September, 1999, pp. 757–761.
[44]N. Petrick, H. P. Chan, B. Sahiner, and M. A. Helvie, "Combined adaptive enhancement and region-growing segmentation of breast masses on digitized mammograms," Med. Phys. **26**, 1642–1654 (1999).

[45]D. Wei, H. P. Chan, N. Petrick, B. Sahiner, M. A. Helvie, D. D. Adler, and M. M. Goodsitt, "False-positive reduction technique for detection of masses on digital mammograms: Global and local multiresolution texture analysis," Med. Phys. **24**, 903–914 (1997).
[46]R. M. Haralick, K. Shanmugam, and I. Dinstein, "Texture features for image classification," IEEE Trans. Syst. Man Cybern. **SMC-3**, 610–621 (1973).
[47]C. E. Metz, B. A. Herman, and J. H. Shen, "Maximum-likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data," Stat. Med. **17**, 1033–1053 (1998).
[48]B. Sahiner, H. P. Chan, N. Petrick, R. F. Wagner, and L. M. Hadjiiski, "Feature selection and classifier performance in computer-aided diagnosis: The effect of finite sample size.," Med. Phys. **27**, 1509–1522 (2000).
[49]D. P. Chakraborty, "Maximum likelihood analysis of free-response receiver operating characteristic (FROC) data," Med. Phys. **16**, 561–568 (1989).
[50]D. P. Chakraborty and K. S. Berbaum, "Observer studies involving detection and localization: Modeling, analysis, and validation," Med. Phys. **31**, 2313–2330 (2004).
[51]J. Wei, B. Sahiner, H. P. Chan, N. Petrick, L. M. Hadjiiski, and M. A. Helvie, "Computer aided diagnosis system for mass detection: Comparison of performance on full-field digital mammograms and digitized film mammograms," RSNA 2003, Chicago, 30 November–5 December 2003, p. 387.

Dual system approach to computer-aided detection of breast masses on mammograms

Jun Wei
Heang-Ping Chan
Berkman Sahiner
Lubomir M. Hadjiiski
Mark A. Helvie
Marilyn A. Roubidoux
Chuan Zhou
Jun Ge

Department of Radiology
University of Michigan, Ann Arbor

Correspondence :
Jun Wei, Ph.D.
Department of Radiology
University of Michigan
CGC B2103
1500 E. Medical Center Drive
Ann Arbor, MI 48109
Phone: 734-647-8553
Fax: 734-615-5513
Email: jvwei@umich.edu
Running head: Dual CAD system for mammographic mass detection

Abstract

In this study, our purpose was to improve the performance of our mass detection system by using a new dual system approach which combines a CAD system optimized with ”average” masses with another CAD system optimized with “subtle” masses. The two single CAD systems have similar image processing steps, which include prescreening, object segmentation, morphological and texture feature extraction, and false positive (FP) reduction by rule-based and linear discriminant analysis (LDA) classifiers. Stepwise feature selection with simplex optimization was applied to each of the single CAD systems during the training. A feed-forward backpropagation artificial neural network (BP-ANN) was trained to merge the scores from the LDA classifiers in the two single CAD systems and differentiate true masses from normal tissue. For an unknown test mammogram, the two single CAD systems are applied to the image in parallel to detect suspicious objects. A total of three data sets were used for training and testing the systems. The first data set of 230 current mammograms, referred as the average data set, was collected from 115 patients. We also collected 264 mammograms, referred as the subtle data set, which were one to two years prior to the current exam from these patients. Both the average and the subtle data sets were partitioned into two independent data sets in a cross validation training and testing scheme. A third data set

containing 65 cases with 260 normal mammograms was used to estimate the false positive (FP) marker rates during testing. The detection performance of the CAD system was assessed by free response receiver operating characteristic (FROC) analysis. When the single CAD system trained on the average data set was applied to the test set, the FP marker rates were 2.2, 1.8 and 1.5 per image at the case-based sensitivities of 90%, 85% and 80%, respectively. With the dual CAD system, the FP marker rates were improved to 1.2, 0.9 and 0.7 per image, respectively, at the same case-based sensitivities. Our results indicate that the dual CAD system approach can improve significantly (p<0.05) the performance of mass detection on mammograms compared to that obtained by training a single CAD system with the average masses alone or with both the average and the subtle masses.

Keywords: computer-aided detection (CAD), mass detection, mammogram, dual system, artificial neural network (ANN)

I. INTRODUCTION

Breast cancer is one of the leading causes of cancer mortality among women[1]. It has been reported that early diagnosis and treatment can significantly improve the chance of survival for patients with breast cancer[2-4]. At present, the most successful method for the early detection of breast cancer is screening mammography[5]. Various methods are being developed to improve the accuracy of breast cancer detection. Double reading by radiologists can reduce the miss rate of radiographic reading. However, double reading will increase the cost of mammographic screening. An alternative method is to use a trained computer-aided detection (CAD) system as a second reader[6,7]. Recent clinical studies have shown that CAD systems are helpful for increasing radiologists' accuracy in detecting breast cancers[8-13].

A large volume of literature has been published in the CAD area. CAD systems for mammography generally consist of two subsystems: one is a mass detection system and the other is a microcalcification detection system. Detection of masses on mammograms is often more challenging than detection of microcalcifications. The mass detection systems to-date employed a single-system approach using various techniques for prescreening of mass candidates and classification of true and false positives[14-24]. Our laboratory incorporated two-view mammographic information for improved

differentiation of true masses and false positives and obtained promising preliminary results.[22] However, development of new methods to improve the performance of mass detection systems remains an important area of CAD research.

The CAD systems developed so far mostly used masses seen on current mammograms (i.e., the mammograms on which the masses were detected by radiologists) for training. An important purpose of a CAD system is that it is used as a second reader to alert radiologists to subtle cancers that may be overlooked. To study the ability of a CAD system in detecting subtle cancers that are likely to be missed by radiologists, one way is to evaluate its accuracy in detecting missed cancers on prior mammograms (i.e., the mammograms in previous examinations on which the mass or cancer can be seen retrospectively but was considered negative or benign at the time of the examination). Some researchers have investigated the performance change of CAD systems when using prior mammograms as input. In our study of mass detection on prior mammograms[25], we obtained a case-based sensitivity of 74% (20/27) of the malignant masses with 2.2 false positives (FPs) per image. te Brake et al[26] reported that their CAD system has a case-based sensitivity of 34% (22/65) of the cancers which have the appearance of masses or stellate lesions in the prior examinations with 1 FP per image. A commercial system (R2 ImageChecker) also reported detection of 42% (72/172) of the cancers in the prior years

which were considered worthy of call-back in retrospect by expert mammographers with about 2 FP marks/case[27]. Zheng et al.[23] reported that their CAD system trained with current mammograms could not perform optimally in prior mammograms and vice versa; whereas the same system trained with prior mammograms can perform better on detecting the masses on prior mammograms. Recently, an assessment study[28] was conducted to compare the performance of two commercial systems and one research CAD system on current mammograms and prior mammograms. The results showed that the true positive (TP) fraction for CAD systems on prior mammograms of 39 breasts with malignant masses ranged from 15% to 26% with 0.28 to 0.41 FP marks/image. Although the detection performance reported in the different studies vary, probably due to the differences in the data set used, these studies indicate that the sensitivity of current CAD systems in detecting subtle masses on prior mammograms are substantially lower than those obtained from detection on current mammograms. The difficulty in recognizing the subtle and possibly different features of the masses on priors compared to those of the masses on current mammograms may be one of the factors that causes oversight for both radiologists and the CAD systems.

The goal of pattern recognition is to achieve the best possible classification performance in the task at hand. Researchers had shown that, for a class of objects with a

wide range of characteristics, the classification performance can be improved by using combination of classifiers whereby objects of certain characteristics are classified by one classifier using a set of features and objects of different characteristics by another classification scheme based on different features[29-35]. The advantage of using combination of classifiers is that it may stabilize the training of classifiers even with a relatively small sample size because each classifier does not have to accommodate a wide range of characteristics and features[36,37]. These observations motivated our interest in the design of a dual CAD system for mass detection.

Since the missed cancers on prior mammograms represent the difficult cases that are more likely to be missed by radiologists if similar cancers occur on screening mammograms, it is important to improve the sensitivity of the CAD system in detecting these cancers. On the other hand, when a CAD system is applied to a new mammogram in clinical practice, it has to detect breast lesions of all degrees of subtlety effectively. However, it is difficult to train a single CAD system to provide optimal detection for all lesions over the entire spectrum of subtlety because the classifiers have to make compromises to accommodate cancers of a wide range of characteristics. Therefore, we have been exploring a new dual CAD system approach that combines a CAD system trained with retrospectively seen masses on prior mammograms with a CAD system

trained with masses detected on current mammograms[38,39]. In this paper, we will describe the design of the dual CAD system and report our current results.

II. MATERIALS AND METHOD

2.1 Data Sets

All mammograms in this study were collected from patient files in the Department of Radiology at the University of Michigan with Institutional Review Board (IRB) approval. The mammograms were digitized with a LUMISYS 85 laser film scanner with a pixel size of 50μm×50μm and 4096 gray levels. The scanner was calibrated to have a linear relationship between gray levels and optical densities (O.D.) from 0.1 to greater than 3 O.D. units. The nominal O.D. range of the scanner is 0–4. The full resolution mammograms were first smoothed with a 2×2 box filter and subsampled by a factor of 2, resulting in 100μm×100μm images. The images at a pixel size of 100μm×100μm were used for the input of our CAD system.

We collected three data sets. The first data set contained 115 cases with confirmed masses. Each case included the current mammograms that prompted the radiologist to work up the mass. This is referred to as the "average" mass set. All of the cases in the average mass set had two mammographic views: the craniocaudal (CC) view

and the mediolateral oblique (MLO) view or the lateral view, thus yielding a total of 230 mammograms. There were 115 masses (67 malignant masses and 48 benign masses) in this data set, of which 105 were biopsy-proven and 10 were determined to be benign by long-term follow-up.

The second data set was composed of the prior mammograms dated one to two years earlier than the mammograms of the same patients in the average mass set. Since the masses on prior mammograms are on average subtler than those on current mammograms, this data set is referred to as the "subtle" mass set. On five of the 115 patients, no mass or focal density could be identified on either view of the prior mammograms. Therefore, the subtle mass set was composed of 110 cases (62 malignant and 48 benign). For the purpose of training the subtle mass detection system, the subtle masses do not have to be obtained from the same cases as the average mass set but we used the available prior mammograms for these mass cases in our database. Nineteen of the 110 cases had two prior mammogram examinations. Of the 129 examinations in the subtle mass set, 123 had two mammographic views and 6 had three views, with a total of 264 mammograms. Many of the subtle masses on the prior mammograms could be identified only as a focal density corresponding to the location of the subsequently detected mass on the current mammograms. On 44 of the two-view prior mammograms,

the mass location was evident only on one view. Table I summarizes the information for the average and subtle mass subsets.

The third data set was composed of 260 normal bilateral two-view mammograms obtained from 65 patients. No masses were evident on these mammograms upon review by the experienced radiologist.

The two mass data sets were used to estimate the detection sensitivity and the normal data set was used for estimating the FP marker rate. For the mass data sets, the true locations of the masses were identified by an experienced MQSA radiologist using all available imaging and clinical information. The radiologist also provided an estimate of the longest diameter of the mass, descriptors of its margin and shape, a visibility rating, and an estimate of the breast density in terms of BI-RADS category. Figure 1 shows the distributions of mass sizes, mass shapes, mass margins, and their visibility on a 10-point rating scale with 1 representing the most visible masses and 10 the most difficult case relative to the cases seen in their clinical practice. The masses had a mean of 13.7 mm and a median of 12 mm in the average data set and a mean of 9.7 mm and a median of 10 mm in the subtle data set. Figure 2 shows the breast density for both the normal data set and the mass data sets. As can be seen from the distributions of the mass characteristics, the average masses on the current mammograms and the subtle masses on the priors had

large overlap. Nevertheless, on average, the subtle masses were smaller in size and less conspicuous on the mammograms.

2.2 Methods

In order to improve the sensitivity of detecting breast lesions of all degrees of subtlety, we developed a new dual system approach which combines a system trained with average masses with another system trained with subtle masses. When the trained dual system is applied to an unknown mammogram, the two CAD systems are used in parallel to detect suspicious objects on a single mammogram. No prior mammogram is needed. The additional FPs from the use of the two systems are reduced by an information fusion stage. We will refer to the two systems separately trained with the average masses and the subtle masses as "single" CAD systems in the following discussions.

We randomly separated the mass data sets by case into two independent subsets. Both the average and subtle mass subsets followed the same case grouping so that mammograms from the same case would not be separated into the training subset for one single CAD system and the test subset for the other single CAD system in a cross-

validation cycle. Table I shows the subsets of cases in the average and subtle mass data sets. Two-fold cross validation was used for training and testing the algorithms. The training included selecting proper parameters for each single CAD system and for information fusion. Once the training with one mass subset was completed, the parameters were fixed for testing with the other mass subset. The training and test mass subsets were switched and the training and test processes were repeated. The CAD systems were trained with single mammograms. To maximize the number of training images with masses, all images with a visible mass were included regardless of whether they were a part of a two-view or three-view case when the subtle mass subset was used as a training set. However, when the subtle mass subset was used as a test set, only two views were included for each case because we used two-view mammograms to derive the case-based test performance. For cases containing three views, we therefore included only two of the views in testing. We also included cases with the mass visible on only one of the two views. After the two-fold cross validation testing, the overall detection performance was evaluated by combining the performances of the two test subsets. The trained algorithms with the fixed parameters were also applied to the normal set of mammograms, which was not used during training, to estimate the FP rate in screening mammograms.

2.2.1 Single CAD system overview

The major steps in the two single mass detection systems are similar but the feature spaces and classifiers for FP reduction in each system were designed separately to suit the characteristics of average and subtle masses, respectively. The two systems are therefore described together below but the differences will be pointed out whenever applicable. Each single CAD system consists of four processing steps: (1) pre-screening of mass candidates, (2) segmentation of suspicious objects, (3) feature extraction and analysis, and (4) FP reduction by classification of normal tissue structures and masses. The block diagram for the detection scheme is shown in Figure 3.

For the pre-screening stage, we have developed a two-stage gradient field analysis method which not only uses the shape information of masses on mammograms but also incorporates the gray level information of the local object segmented by a region growing technique in the second stage to refine the gradient field analysis[24,40]. Locations of high radial gradient convergence are labeled as mass candidates. After prescreening, the suspicious objects are identified by using a two-stage segmentation method[41]. First, the background-corrected ROI is weighted by a Gaussian function with σ=256 pixels. Then, a k-means clustering using the pixel values in a background-corrected image and a

Sobel filtered image as features is used to segment the object. For each suspicious object, eleven morphological features[21] were extracted. Rule-based and linear discriminant classifiers were trained by using the training data set only to remove the detected structures that were substantially different from breast masses. For the system trained with average masses, global and local multi-resolution texture analysis[42] were performed in each ROI by using the spatial gray level dependence (SGLD) matrices. A total of 364 features were extracted from global texture analysis. Local texture features were extracted from the local region containing the detected object and the peripheral regions within each ROI. A total of 208 features were extracted for local texture analysis. For the system trained with subtle masses, instead of the SGLD texture features, gray level features and run length statistics analysis (RLS) texture features[43] were extracted inside and outside of each mass region on the original image and gradient field image. The gray level features included the contrast of the object relative to the surrounding background, the minimum and the maximum gray levels, and the characteristics derived from the gray level histogram in the regions inside and outside of each object including skewness, kurtosis, energy, and entropy. Five RLS texture features were extracted in both the horizontal and vertical directions: short runs emphasis, long runs emphasis, gray level nonuniformity, run length nonuniformity and run percentage. A total of 66 features were

extracted for the system trained with subtle masses.

In order to obtain the best texture feature subset and also reduce the dimensionality of the feature space to design an effective classifier, stepwise feature selection with linear discriminant analysis (LDA) was applied to the training subset. The detailed procedure has been described elsewhere[24,44,45]. Briefly, at each step one feature was entered or removed from the feature pool by analyzing its effect on the selection criterion, which was chosen to be the Wilks' lambda in this study. Since the appropriate values of thresholds for feature entry, feature elimination, and tolerance of correlation for feature selection were unknown, we used an automated simplex optimization method to search for the best combination of thresholds in the parameter space. The simplex algorithm used a leave-one-case-out resampling method within the training subset to select features and estimate the weights for the LDA classifier. To have a figure-of-merit to guide feature selection, the test discriminant scores from the left-out cases were analyzed using receiver operating characteristic (ROC) methodology[46]. The accuracy for classification of masses and FPs was evaluated as the area under the ROC curve, A_z. In this approach, feature selection was performed without the left-out case so that the test performance would be less optimistically biased[47]. However, the selected feature set in each leave-one-case-out cycle could be slightly different because every cycle had one

training case different from the other cycles. In order to obtain a single trained classifier to apply to the independent test subset, a final stepwise feature selection was performed with the best combination of thresholds, found in the simplex optimization procedure, on the entire training subset to obtain the final set of features and estimate the weights of the LDA. Note that the entire process of feature selection and classifier weight estimation was performed within the training subset. The LDA classifier with the selected feature set was then fixed and applied to the independent test subset. The training and testing processes were performed independently for the two-fold cross-validation sets.

2.2.2 Training and test for dual system

The block diagram for the dual system is shown in Figure 4. During the training of the dual system, we used the current and prior mammograms from the same patients. The current mammograms that contained the average masses were only used to train the first single CAD system. The prior mammograms that contained the subtle masses were only used to train the second single CAD system. The prescreening and the segmentation steps in the two systems are identical. Since the morphological appearances of average and subtle masses are different, the rules in the morphological rule-based FP classification are trained differently for the two single CAD systems. During testing with an independent mammogram, the dual system keeps all the suspicious objects that satisfy

the FP classification rules of either single CAD system and applies the LDA classifiers from both single systems to each object. Each object thus has two LDA scores.

To merge the information from the two CAD systems, a fusion scheme was developed for our dual system. In this study, a feed-forward backpropagation artificial neural network (BP-ANN) was trained to classify the masses from normal tissues by combining the output information from the two single CAD systems. The LDA classifiers from the two single CAD systems were applied to each detected object . The two LDA discriminant scores for each object were used as input to the BP-ANN. The BP-ANN had an input layer with two nodes, a hidden layer with N nodes, and an output layer with one node. The nodes were interconnected by weights and information propagated from one layer to the next through a log-sigmoidal activation function. The learning of the ANN was a supervised process in which known training cases were input to the ANN. The performance function for the network was the mean-squared error between the network outputs and the target outputs. The weights of the network were adjusted iteratively by a feedforward backpropagation procedure to minimize the error. Detailed description of the backpropagation neural network can be found in the literature[48,49].

To choose the number of hidden nodes (N) in the BP-ANN, we used a 3-fold cross-validation method within the training subset. We randomly separated the entire training subset including all detected objects into three independent groups. The objects belonging to the same case were separated into the same group. For a given N, three training cycles were performed, in each of which two of the three groups were used to train the BP-ANN and the left-out group was used to test its performance. The A_z value obtained from the ANN output scores for the test group was used as the performance index for that training cycle. The average of the A_z values from the three test groups represented the performance of the BP-ANN with N hidden nodes. In our experiment, a BP-ANN with 3 hidden nodes provided the largest average A_z value and was therefore chosen. The weights of the chosen BP-ANN were retrained with the entire training subset. The BP-ANN with the trained weights was used to merge the information from the two single CAD systems.

To test the dual system, the two trained single CAD systems, one trained with the average mass set and the other with the subtle mass set, were applied in parallel to each single "unknown" mammogram in the independent test subset. No prior mammogram was needed during testing.

2.2.3 Evaluation methods

The detected individual objects were compared with the "truth" ROI marked by the experienced radiologist, as described above. A detected object was scored as TP if the overlap between the bounding box of the detected object and the bounding box of the true mass relative to the larger of the two bounding boxes was over 25%. Otherwise, it would be scored as FP. The 25% threshold was selected as described in our previous study[21].

The FP marker rate was estimated in two ways: one from detection on the same test subsets with masses, the other from detection on the normal data set of negative mammograms. For the latter, we applied the trained dual CAD system to the normal data set. The number of FP marks produced by the CAD system was determined by counting the detected objects on the normal cases. The mass detection sensitivity was determined by counting the detected masses on the test mass subset. The detection performance of the CAD system was assessed by free response ROC (FROC) analysis. An FROC curve was obtained by plotting the mass detection sensitivity as a function of FP marks per image either obtained from the mass data subset or the normal set at the corresponding

decision threshold.

FROC curves were presented on a per-mammogram and a per-case basis. For image-based FROC analysis, the mass on each mammogram was considered an independent true object. For case-based FROC analysis, the same mass imaged on the two-view mammograms was considered to be one true object and detection of either or both masses on the two views was considered to be a TP detection.

Since we used two-fold cross validation method for training and testing, we obtained two test FROC curves, one for each test subset, for each of the conditions (e.g., single CAD system approach or dual system approach). To summarize the results for comparison, an average test FROC curve was derived by averaging the FP rates at the same sensitivity along the FROC curves of the two corresponding test subsets.

In order to compare the performance of the single CAD system and the dual CAD system, we applied the alternative free-response ROC (AFROC) method and the jackknife free-response ROC (JAFROC) method developed by Chakraborty et al.[50,51] to the pairs of FROC curves. In the AFROC method, the FROC data are first transformed by counting the number of false-positive images (FPI) instead of the FPs per image. The confidence rating of an FPI is determined by the highest confidence FP decision on the image regardless of how many lower confidence FP decisions are made on the same

image. The ROCKIT curve fitting software and statistical significance tests for ROC analysis developed by Metz et al.[46] can then be used to analyze the AFROC data.

III. RESULTS

Figure 5 shows an example of the 2-dimensional feature space that was used as the input to the BP-ANN being trained to merge the information from the two single CAD subsystems. The two features are the output scores of the LDA classifiers trained with the average masses and with the subtle masses. The correlation coefficients of the two features are 0.46 and 0.44 for each of the training subsets, respectively. The low correlation indicated that the two single CAD systems extracted relatively independent features from the object. The validation A_z values of the chosen ANN on the two training subsets were 0.92±0.01 and 0.87±0.01, respectively. The ANN classifiers achieved A_z values of 0.90±0.02 and 0.89±0.01 on the two independent test subsets, respectively. Figure 6 shows the ROC curves for the two test subsets.

In order to evaluate the effectiveness of our dual system approach, we compared its performance on the test subsets containing average masses with two other single CAD systems: the CAD system trained only on the average mass set and the CAD system trained on both the average and the subtle mass sets. When a single CAD system was

trained only with the average masses, the number of selected features was 21 (14 global and 7 local) and 16 (10 global and 6 local) texture features for the two independent training subsets, respectively. When the CAD system was trained with both the average and the subtle masses, the number of selected features was 17 (11 global and 6 local) and 18 (7 global and 11 local) texture features for the two independent training subsets, respectively.

For the dual system, the single system trained with the average masses was the same as that described above. For the single system trained with subtle masses, four (2 gray level and 2 RLS texture) and five (3 gray level and 2 RLS texture) features were selected for the two independent training subsets, respectively.

The average test FROC curves of the dual CAD system on the test subsets with average masses were compared to those of the single CAD systems in Figure 7. The FP rates were estimated from the mass data set. The dual CAD system achieved a case-based sensitivity of 80%, 85%, and 90% at 0.6, 0.8, and 1.0 FPs/image, respectively, compared with 1.3, 1.5, and 1.8 FPs/image on the single CAD system trained with average masses alone. The performance of the single CAD system trained with both the average masses and the subtle masses was comparable to that trained with average masses alone, with FP rates of 1.4, 1.6, and 1.8 FPs/image at the same sensitivities,

respectively. Figure 8 shows the comparison of the three average test FROC curves, similar to those shown in Figure 7, except that the FP rates were estimated from the normal data set. The FP rates at a few selected sensitivities for the dual and single CAD systems were summarized in Table II.

In this study, we have 67 malignant cases in the average mass set. Figure 9 compares the average test FROC curves of the single CAD system and the dual system for detection of malignant masses. The result for the single CAD system trained with average masses was shown and the FP rate was estimated from the mammograms without masses. In this case, the dual CAD system achieved a case-based sensitivity of 80%, 85%, and 90% at 0.6, 0.9, and 1.2 FP marks/image, respectively, compared with 1.1, 1.6, and 2.0 FP marks/image on the single CAD system.

An important purpose of a CAD system is to serve as a second reader to alert radiologists to subtle cancers that may be overlooked. Figure 10 and Figure 11 compare the average FROC curves of the single CAD system and the dual system for detection in the test subsets with subtle masses. The TP rate in Figure 10 was estimated by including both malignant and benign masses and that in Figure 11 was estimated from malignant masses only. The single CAD system trained with average masses alone was used. The FP rates for both systems were estimated from the mammograms without masses. The

dual CAD system achieved a case-based sensitivity of 50% at 0.7 FP marks/image for all masses and at 0.5 FP marks/image for malignant masses only, compared with 1.4 FP marks/image for all masses and 1.1 FP marks/image for malignant masses only using the single CAD system.

Table II summaries the test results on the average and subtle mass sets for the dual system and the single CAD system trained with average masses at different sensitivity levels. The FP marker rates were estimated from the detection on the normal data set.

IV. DISCUSSION

The masses on prior mammograms are subtler and more difficult to detect than the masses on current mammograms. In this study, we developed a dual CAD system, which combines a system trained with masses on prior mammograms and a system trained with masses detected on current mammograms. We have demonstrated that this dual system can increase the accuracy of detecting both average masses and subtle masses. The comparisons of the dual system with that of the single CAD system trained with average masses alone and that of the single CAD system trained with both average and subtle masses (Figure 7) indicate that the gain in the detection accuracy of the dual

system could not be achieved by simply using a larger training set with both average and subtle masses. In fact, it is interesting to note that the performance of the single CAD system trained with both the average and the subtle masses appeared to be degraded slightly, in comparison with the single system trained with average masses alone, when it was applied to the test set of average masses. The decreased performance may reflect the compromise made when the single CAD system was trained to accommodate a wide range of lesion characteristics. Thus, the dual system approach may have improved its performance through other factors, including the flexibility in using different feature spaces and training the parameters for each type of masses and the information fusion combining the two single CAD systems effectively.

For the comparison of the different systems, we analyzed the false negatives (FN) of the single CAD systems and the dual CAD system when the test subsets with average masses were used. It was found that the FN rates of the single CAD system trained with average masses, the single CAD system trained with subtle masses, and the dual system were 23.9% (55/230), 28.3% (65/230) and 16.5% (38/230), respectively, after FP reduction by the morphological LDA classifier in each system. Twenty-nine masses were missed by both of the single systems. By using the dual system, 53 masses that were FNs for either single system could be detected. However, the masses that were missed by

both of the single CAD systems could not be recovered by the dual CAD system.

The comparison of the FROC curves for the dual CAD system and the single CAD system in terms of the area under the fitted AFROC curve (A_1) and the p values for both test subsets with average masses was summarized in Table III. The differences between the A_1 values for the two systems were statistically significant ($p<0.05$). The fitted AFROC curves, however, did not fit very well to the transformed AFROC data, as we discussed previously[24]. For the JAFROC method, Chakraborty et al. provided software to estimate the statistical significance of the difference between two FROC curves. The comparison of the figure-of-merit (FOM) and the p values was also summarized in Table III. The differences between the FOM of the dual CAD system and that of the single CAD system for both test subsets were again statistically significant ($p<0.05$).

The performance of the dual system in detecting subtle masses was also superior to that of the single system trained with average masses (Figure 10). To analyze these results statistically, JAFROC and AFROC methods were also used. It was found that the differences between the results of the dual CAD system and those of the single CAD system on the two test subsets containing subtle masses were statistically significant by

both JAFROC and AFROC methods. The comparison of the area under the fitted AFROC curve (A_1), the FOM, and the p values was summarized in Table IV.

Our motivation of this study is to improve the performance of a CAD system for mass detection. A CAD detection system is generally intended for use in screening mammography. At the screening stage, all lesions of concern should be pointed out to radiologists so that the radiologists can judge if a recall is warranted. If a detection system is trained to mark only the malignant lesions, it may be attempting to play the role of a triage system (alerting radiologists to work up only "malignant" cases) rather than that of a second reader. Furthermore, since computerized lesion detection or characterization on mammograms is not 100% sensitive, it will be confusing to the radiologists whether an unmarked suspicious lesion is missed or it is considered benign by the computer. We believe that computer-aided diagnosis (CADx) may be used in different ways in conjunction with a CAD detection system, for example, the likelihood of malignancy may be estimated by the CADx system and displayed for every detected lesion, and/or a CADx system may be used during diagnostic workup. Either way the CAD system will first alert radiologists to all masses, leaving the assessment of malignancy or benignity to a second stage. The training set thus included both malignant and benign masses.

For a CAD system, its performance for detecting malignant masses is more important than its performance for detecting all masses. The FROC curves for detection of malignant masses on the average data set and the subtle data set, shown in Figure 9 and 11, respectively, indicated that the dual system could also achieve an improvement in the detection performance over that of the single system. The differences in the A_1 and the FOM for the detection of malignant cases in the average and subtle mass test subsets were statistically significant, as shown in Table III and IV, respectively.

In screening mammography, the cancer rate is 3 to 5 per 1000. Most of the mammograms are normal. Therefore, some CAD researchers and users estimate the FP rate using normal mammograms[52-54] because it reflects how the CAD system performs in terms of specificity and whether the CAD system may cause extra efforts for radiologists to double check the marked locations or unnecessary recalls in a screening setting. Furthermore, for CAD systems that set a maximum number of detected objects at the output, estimating the number of FPs using images with lesions can potentially lead to an optimistic bias for the FROC curve because one of the detected objects will likely be the true lesion. The FP rate can thus be underestimated by as much as 1 per image. In addition, the JAFROC analysis requires that the FP rates be estimated on normal images. We therefore reported the FP rates of our CAD systems on both mammograms with

masses and without masses to facilitate comparison with other CAD systems in case investigators may evaluate their FP rates in either way.

In this study, we evaluated the performance of the trained CAD systems with an independent test set using the 2-fold cross validation method. Although the selection of parameters and features was performed using the training set, we had the full knowledge of the performance for the test set so that the selections could be optimistically biased. True independent testing will have to be performed with unknown cases that have never been used for testing the CAD system before, such as those in a prospective clinical trial. However, this test step is beyond the scope of our current developmental process. Since we used the same cross-validation method for evaluated of the dual system and the single CAD systems, the comparison of their relative performances is expected to be less biased than their individual performances.

V. CONCLUSION

We have proposed a new dual system approach which combines a system trained with subtle masses on prior mammograms and a system trained with average masses on current mammograms. The dual system achieved higher sensitivities at the corresponding FP rates than a single CAD system trained with average masses alone or

trained with both average masses and subtle masses. Alternatively, the dual system had lower FP rates than the single CAD system at corresponding sensitivities. The improvement in the FROC curves by the dual system approach was found to be statistically significant ($p<0.05$) for both average masses and subtle masses using either the AFROC or the JAFROC method. Our results indicate that the dual system approach is promising for improving the performance of CAD systems for mass detection on mammograms.

Acknowledgments

This work is supported by U. S. Army Medical Research and Materiel Command grants W81XWH-1-04-1-0475, DAMD 17-02-1-0214, and USPHS grant CA95153. The content of this paper does not necessarily reflect the position of the government and no official endorsement of any equipment and product of any companies mentioned should be inferred. The authors are grateful to Charles E. Metz, Ph.D., for the LABROC program and to Dev Chakraborty, Ph.D., for the JAFROC program.

Table I. Description of cases in the average and subtle mass data sets and the subsets for training and testing in the cross-validation scheme.

	Mass Subset 1		Mass Subset 2	
	Average mass subset	Subtle mass subset	Average mass subset	Subtle mass subset
Total no. of cases	57	54	58	56
Cases with two prior examinations	NA	10	NA	9
Exams with two-views	57	58	58	65
Exams with three-views	0	6	0	0
Total no. of images	114	134	116	130
No. of negative images	0	25	0	19
No. of mass images for training	114	109	116	111
No. of two-view pairs for testing	57	64	58	65
No. of images for testing	114	128	116	130
No. of malignant masses	36	33	31	29
No. of benign masses	21	21	27	27

Table II. Comparison of case-based detection performance between the dual system and the single CAD system trained with average masses alone. The FP marker rates were estimated from detection on the normal data set. The FROC curves were obtained by averaging the FROC curves of the two test subsets.

TP	Average mass test set (FP marks/image)		Subtle mass test set (FP marks/image)	
	Single system	Dual system	Single system	Dual system
90%	2.2	1.2		
80%	1.5	0.7		2.8
70%	1.0	0.3	2.4	2.3
60%	0.5	0.2	1.8	1.5
50%	0.3	0.1	1.4	0.7

Table III. Estimation of the statistical significance in the difference between the FROC performance of the dual system and the single CAD system trained with average masses alone when the systems were evaluated on the average mass test subsets. The FROC curves with the FP marker rates obtained from the normal data set were compared.

	A_1 (AFROC)				FOM (JAFROC)			
	All Cases		Malignant Cases		All Cases		Malignant Cases	
	Test subset 1	Test subset 2	Test subset 1	Test subset 2	Test subset 1	Test subset 2	Test subset 1	Test subset 2
Single system	0.45	0.44	0.47	0.52	0.48	0.48	0.53	0.55
Dual system	0.55	0.53	0.58	0.62	0.60	0.56	0.63	0.64
P values	0.0004	0.0156	0.0003	0.0318	<0.0001	0.007	0.0004	0.0252

Table IV. Estimation of the statistical significance in the difference between the FROC performance of the dual system and the single CAD system trained with average masses alone when the systems were evaluated on the subtle mass test subsets. The FROC curves with the FP marker rates obtained from the normal data set were compared.

	A_1 (AFROC)				FOM (JAFROC)			
	All Cases		Malignant Cases		All Cases		Malignant Cases	
	Test subset 1	Test subset 2	Test subset 1	Test subset 2	Test subset 1	Test subset 2	Test subset 1	Test subset 2
Single system	0.17	0.20	0.24	0.25	0.21	0.23	0.24	0.26
Dual system	0.28	0.25	0.35	0.34	0.30	0.28	0.36	0.34
P values	<0.0001	0.046	<0.0001	0.0067	0.0007	0.048	<0.0001	0.0035

References

[1]"American Cancer Society, www.cancer.org 2005, "Statistics for 2005"," (2005).

[2]C. R. Smart, R. E. Hendrick, J. H. Rutledge, and R. A. Smith, "Benefit of mammography screening in women ages 40 to 49 years: current evidence from randomized controlled trials," Cancer 75, 1619-1626 (1995).

[3]S. A. Feig, C. J. D'Orsi, R. E. Hendrick, V. P. Jackson, D. B. Kopans, B. Monsees, E. A. Sickles, C. B. Stelling, M. Zinninger, and P. Wilcox-Buchalla, "American College of Radiology guidelines for breast cancer screening," AJR Am J Roentgenol. 171, 29-33 (1998).

[4]B. Cady and J. S. Michaelson, "The life-sparing potential of mammographic screening," Cancer 91, 1699-1703 (2001).

[5]H. C. Zuckerman, *The role of mammography in the diagnosis of breast cancer. In: Breast Cancer, Diagnosis and Treatment*, I. M. Ariel and J. B. Cleary, (McGraw-Hill, New York, 1987).

[6]F. Shtern, C. Stelling, B. Goldberg, and R. Hawkins, "Novel technologies in breast imaging: National Cancer Institute perspective," Orlando, Florida,

[7]C. J. Vyborny, "Can computers help radiologists read mammograms?," Radiology 191, 315-317 (1994).

[8]T. W. Freer and M. J. Ulissey, "Screening mammography with computer-aided detection: Prospective study of 12,860 patients in a community breast center," Radiology 220, 781-786 (2001).

[9]M. A. Helvie, L. M. Hadjiiski, E. Makariou, H. P. Chan, N. Petrick, B. Sahiner, S. C. B. Lo, M. Freedman, D. Adler, J. Bailey, et al., "Sensitivity of noncommercial computer-

aided detection system for mammographic breast cancer detection - A pilot clinical trial," Radiology 231, 208-214 (2004).

[10]R. L. Birdwell, P. Bandodkar, and D. M. Ikeda, "Computer-aided Detection with Screening Mammography in a University Hospital Setting," Radiology 236, 451-457 (2005).

[11]D. Gur, J. H. Sumkin, H. E. Rockette, M. A. Ganott, C. Hakim, L. A. Hardesty, W. R. Poller, R. Shah, and L. Wallace, "Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system," J National Cancer Institute 96, 185-190 (2004).

[12]S. A. Feig, E. A. Sickles, W. P. Evans, and M. N. Linver, "Re. Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system," J National Cancer Institute 96, 1260-1261 (2004).

[13]T. E. Cupples, "Impact of computer-aided detection (CAD) in a regional screening mammography program," Radiology 221(P), 520 (2001).

[14]S. L. Ng and W. F. Bischof, "Automated detection and classification of breast tumors," Comput. Biomed. Res. 25, 218-237 (1992).

[15]N. Petrick, H. P. Chan, D. Wei, B. Sahiner, M. A. Helvie, and D. D. Adler, "Automated detection of breast masses on mammograms using adaptive contrast enhancement and texture classification," Med. Phys. 23, 1685-1696 (1996).

[16]B. Zheng, Y. H. Chang, and D. Gur, "Computerized detection of masses in digitized mammograms using single-image segmentation and a multilayer topographic feature analysis," Acad. Radiol. 2, 959-966 (1995).

[17]N. Karssemeijer and G. te Brake, "Detection of stellate distortions in mammograms," IEEE Trans. Med. Img. 15, 611-619 (1996).

[18]H. Kobatake and Y. Yoshinaga, "Detection of spicules on mammogram based on skeleton analysis," IEEE Trans Med Img 15, 235-245 (1996).

[19]Z. M. Huo, M. L. Giger, C. J. Vyborny, D. E. Wolverton, R. A. Schmidt, and K. Doi, "Automated computerized classification of malignant and benign masses on digitized mammograms," Acad. Radiol. 5, 155-168 (1998).

[20]W. Qian, L. H. Li, and L. P. Clarke, "Image feature extraction for mass detection in digital mammography: Influence of wavelet analysis," Medical Physics 26, 402-408 (1999).

[21]N. Petrick, H. P. Chan, B. Sahiner, M. A. Helvie, S. Paquerault, and L. M. Hadjiiski, "Breast cancer detection: Evalution of a mass detection algorithm for computer-aided diagnosis: Experience in 263 patients.," Radiology 224, 217-224 (2002).

[22]S. Paquerault, N. Petrick, H. P. Chan, B. Sahiner, and M. A. Helvie, "Improvement of computerized mass detection on mammograms: Fusion of two-view information," Med. Phys. 29, 238-247 (2002).

[23]B. Zheng, W. F. Good, D. R. Armfield, C. Cohen, T. Hertzberg, J. H. Sumkin, and D. Gur, "Performance change of mammographic CAD schemes optimized with most-recent and prior image databases," Acad. Radiol. 10, 283-288 (2003).

[24]J. Wei, B. Sahiner, L. M. Hadjiiski, H. P. Chan, N. Petrick, M. A. Helvie, M. A. Roubidoux, J. Ge, and C. Zhou, "Computer aided detection of breast masses on full field digital mammograms," Med. Phys. 32, 2827-2838 (2005).

[25]N. Petrick, H. P. Chan, B. Sahiner, M. A. Helvie, and S. Paquerault, "Evaluation of an automated computer-aided diagnosis system for the detection of masses on prior mammograms," San Diego, 2000.

[26]G. M. Te Brake, N. Karssemeijer, and J. Hendriks, "Automated detection of breast carcinomas not detected in a screening program," Radiology 207, 465-471 (1998).

[27]D. M. Ikeda, R. L. Birdwell, K. F. O'Shaughnessy, E. A. Sickles, and R. J. Brenner, "Computer-aided detection output on 172 subtle findings on normal mammograms previously obtained in women with breast cancer detected at follow-up screening mammography," Radiology 230, 811-819 (2004).

[28]D. Gur, J. S. Stalder, L. A. Hardesty, B. Zheng, J. H. Sumkin, D. M. Chough, B. E. Shindel, and H. E. Rockette, "Computer-aided Detection Performance in Mammographic Examination of Masses: Assessment," Radiology 233, 418-423 (2004).

[29]H. El-Shishini, M. S. Abdel-mottaleb, M. El-Raey, and A. Shoukry, "A multistage algorithm for fast classification of patterns," Pattern Recog. Letters 10, 211-215 (1989).

[30]J. Y. Zhou and T. Pavlidis, "Discrimination of characters by a multi-stage recognition process," Patt. Recog. 27, 1539-1549 (1994).

[31]L. K. Hansen and P. Salamon, "Neural network ensembles," IEEE Trans Pattern Analysis and Machine Intelligence 12, 993-1001 (1990).

[32]G. Rogova, "Combining the results of several neural network classifiers," Neural Networks 7, 777-781 (1994).

[33]S. Hashem and B. Schmeiser, "Improving model accuracy using optimal linear combinations of trained neural networks," IEEE Trans Neural Networks 6, 792-794 (1995).

[34]T. K. Ho, J. J. Hull, and S. N. Srihari, "Decision combination in multiple classifier systems," IEEE Trans Pattern Analysis and Machine Intelligence 16, 66-75 (1994).

[35]D. A. Bell, J. W. Guan, and Y. Bi, "On combining classifier mass functions for text categorization," IEEE Trans on Knowledge and data engineering 17, 1307-1319 (2005).

[36]J. Cao, M. Ahmadi, and M. Shridhar, "Recognition of handwritten numerals with multiple feature and multistage classifier," Patt. Recog. 28, 153-160 (1995).

[37]A. Al-Ani and M. Deriche, "A new technique for combining multiple classifiers using the dempster-shafer theory of evidence," Journal of Artificial Intelligence Research 17, 333-361 (2002).

[38]J. Wei, B. Sahiner, L. M. Hadjiiski, H. P. Chan, H. M. A, and R. M. A, "A dual computer-aided detection (CAD) system for improvement of mass detection on mammograms," RSNA Program Book 2004 491 (2004).

[39]J. Wei, B. Sahiner, L. M. Hadjiiski, H. P. Chan, H. M. A, R. M. A, N. Petrick, C. Zhou, and J. Ge, "Computer aided detection of breast masses on mammograms: performance improvement using a dual system," Proc SPIE Int. Soc. Opt. Eng. 5747, 9-15 (2005).

[40]J. Wei, B. Sahiner, L. M. Hadjiiski, H. P. Chan, N. Petrick, M. A. Helvie, C. Zhou, and Z. Ge, "Computer aided detection of breast masses on full-field digital mammograms: false positive reduction using gradient field analysis," Proc. SPIE Medical Imaging 5370, 992-998 (2004).

[41]N. Petrick, H. P. Chan, B. Sahiner, and M. A. Helvie, "Combined adaptive enhancement and region-growing segmentation of breast masses on digitized mammograms," Med. Phys. 26, 1642-1654 (1999).

[42]D. Wei, H. P. Chan, N. Petrick, B. Sahiner, M. A. Helvie, D. D. Adler, and M. M. Goodsitt, "False-positive reduction technique for detection of masses on digital mammograms: global and local multiresolution texture analysis," Med. Phys. 24, 903-914 (1997).

[43]M. M. Galloway, "Texture classification using gray level run lengths," Comp. Graph. Img Proc. 4, 172-179 (1975).

[44]M. J. Norusis, *SPSS for Windows Release 6 Professional Statistics*, (SPSS Inc., Chicago, IL, 1993).

[45]L. M. Hadjiiski, B. Sahiner, H. P. Chan, N. Petrick, M. A. Helvie, and M. N. Gurcan, "Analysis of Temporal Change of Mammographic Features: Computer-Aided Classification of Malignant and Benign Breast Masses," Med. Phys. 28, 2309-2317 (2001).

[46]C. E. Metz, B. A. Herman, and J. H. Shen, "Maximum-likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data," Stat. Med. 17, 1033-1053 (1998).

[47]B. Sahiner, H. P. Chan, N. Petrick, R. F. Wagner, and L. M. Hadjiiski, "Feature selection and classifier performance in computer-aided diagnosis: The effect of finite sample size.," Med. Phys. 27, 1509-1522 (2000).

[48]C. M. Bishop, *Neural Networks for Pattern Recognition*, (Clarendon Press, Oxford, 1995).

[49]J. A. Freeman and D. M. Skapura, *Neural Networks-Algorithms, Applications, and Programming Techniques*, (Addison-Wesley, Reading, 1991).

[50]D. P. Chakraborty and L. H. L. Winter, "Free-response methodology: Alternate analysis and a new observer-performance experiment," Radiology 174, 873-881 (1990).

[51]D. P. Chakraborty and K. S. Berbaum, "Observer studies involving detection and localization: modeling, analysis, and validation," Medical Physics 31, 2313-2330 (2004).

[52]K. F. O'Shaughnessy, R. A. Castellino, S. L. Muller, and K. Benali, "Computer-aided detection (CAD) on 90 biopsy-proven breast cancer cases acquired on a full-field digital mammography (FFDM) system," Radiology 221(P), 471 (2001).

[53]L. J. Warren Burhenne, S. A. Wood, C. J. D'Orsi, S. A. Feig, D. B. Kopans, K. F. O'Shaughnessy, E. A. Sickles, L. Tabar, C. J. Vyborny, and R. A. Castellino, "Potential contribution of computer-aided detection to the sensitivity of screening mammography," Radiology 215, 554-562 (2000).

[54]R. E. Brem, J. W. Hoffmeister, J. A. Rapelyea, G. Zisman, K. Mohtashemi, G. Jindal, M. P. DiSimio, and S. K. Rogers, "Impact of breast density on computer-aided detection for breast cancer," AJR Am J Roentgenol. 184, 439-444 (2005).

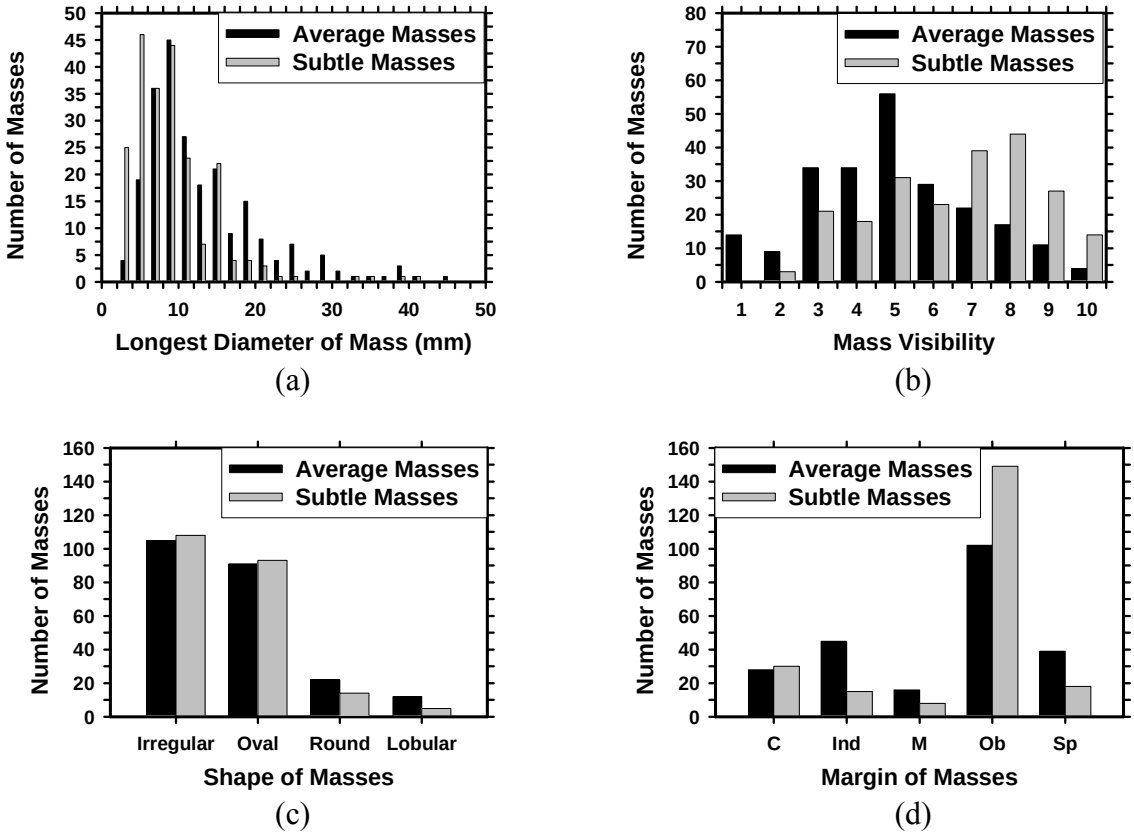


Figure 1. The characteristics of the masses in our mass data set: (a) distribution of mass sizes, (b) distribution of mass visibility on a 10-point rating scale with 1 representing the most visible masses and 10 the most subtle masses relative to the cases seen in clinical practice, (c) distribution of mass shapes, (d) distribution of mass margins, C: circumscribed, Ind: indistinct, M: microlobulated, Ob: obscured, Sp: spiculated.

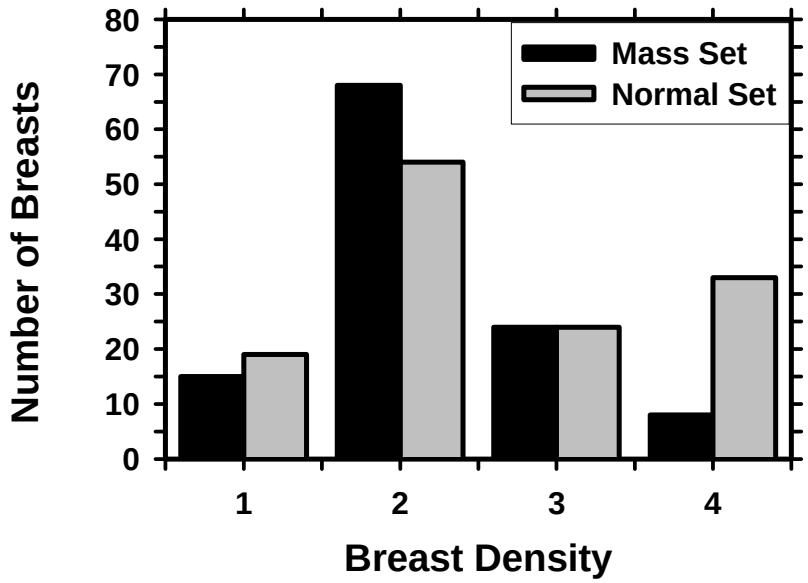


Figure 2. The distribution of breast density in terms of BI-RADS categories estimated by an MQSA radiologist.

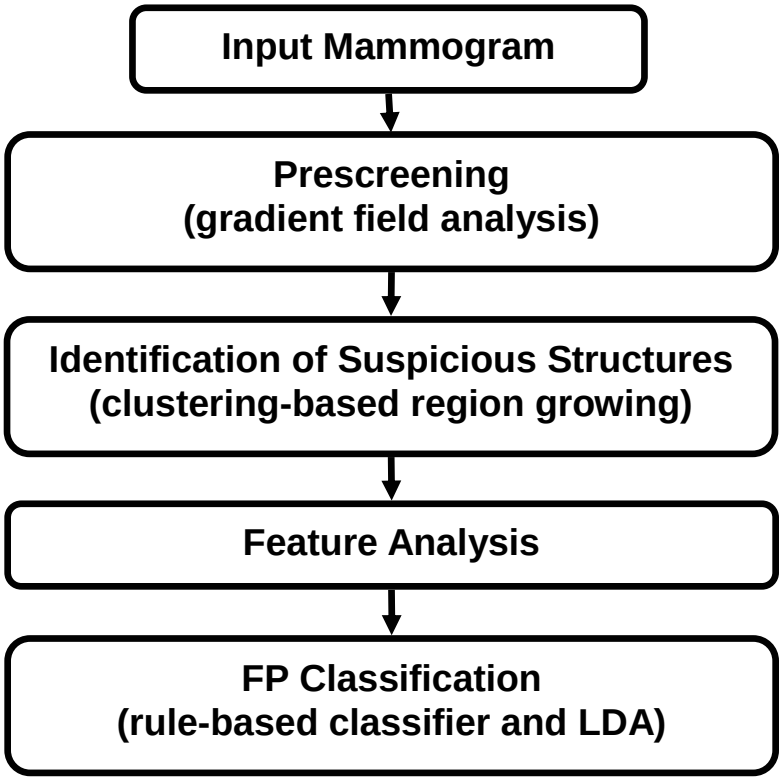


Figure 3. Schematic diagram of our single CAD system for mass detection. The FP classification stage includes rule-based classification, a morphological LDA classifier, and a texture feature LDA classifier for differentiating masses from normal breast tissues.

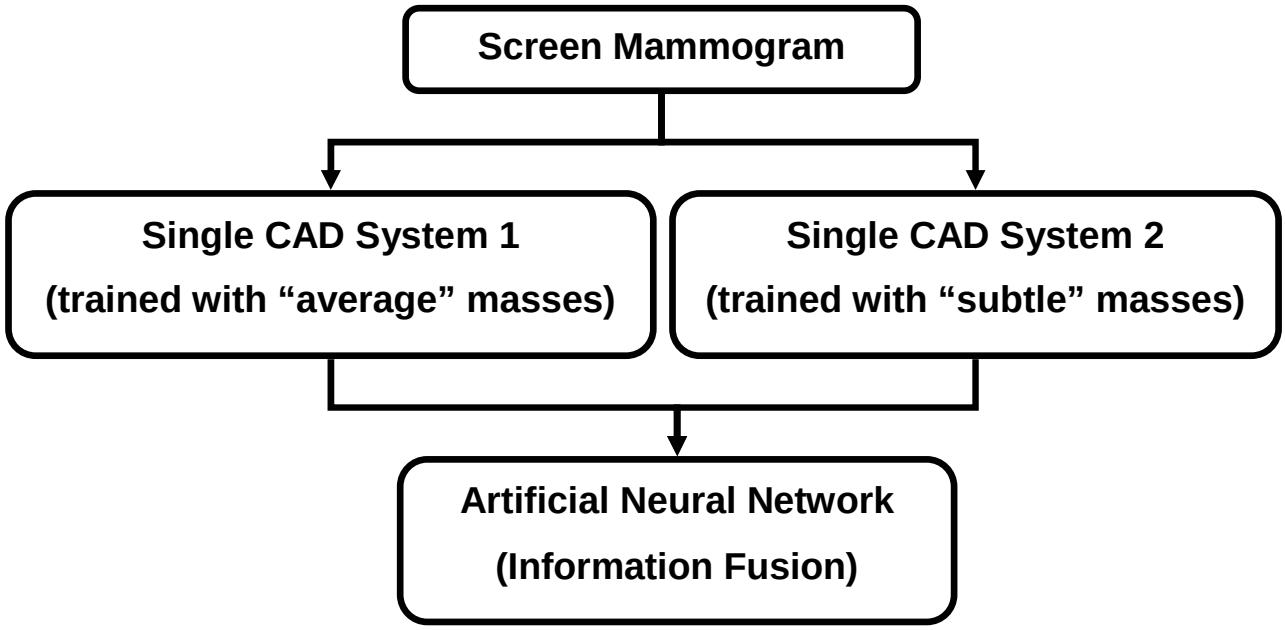


Figure 4. Schematic diagram of proposed dual CAD system for mass detection. BP-ANN is used for information fusion.

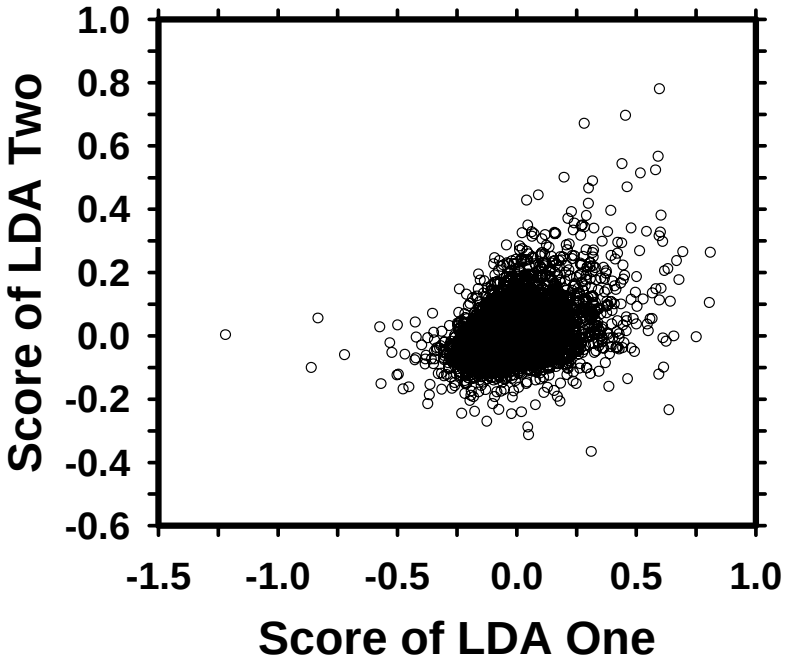


Figure 5. An example of a scatter plot of the LDA scores from the two single CAD systems which are used as input to the BP-ANN. The correlation coefficient between the scores of two LDA classifiers is 0.46, indicating that the two LDA scores are essentially independent features.

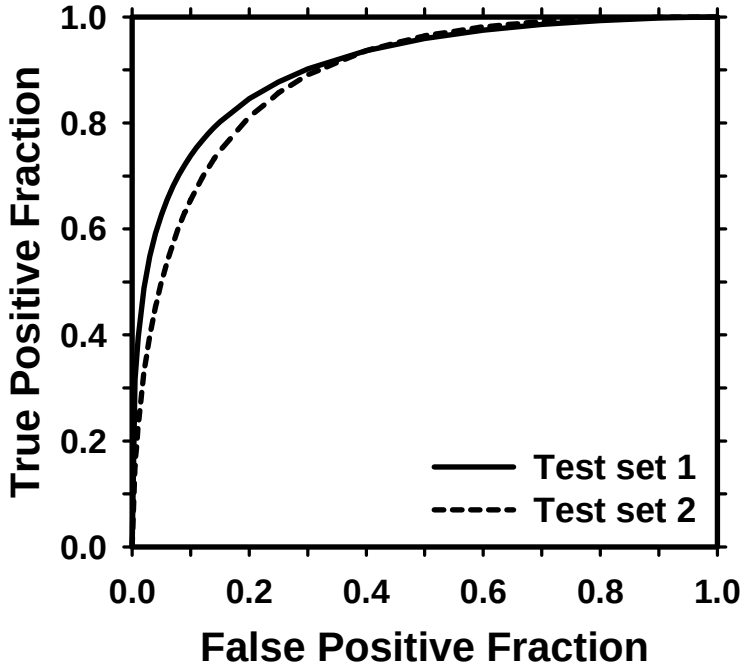
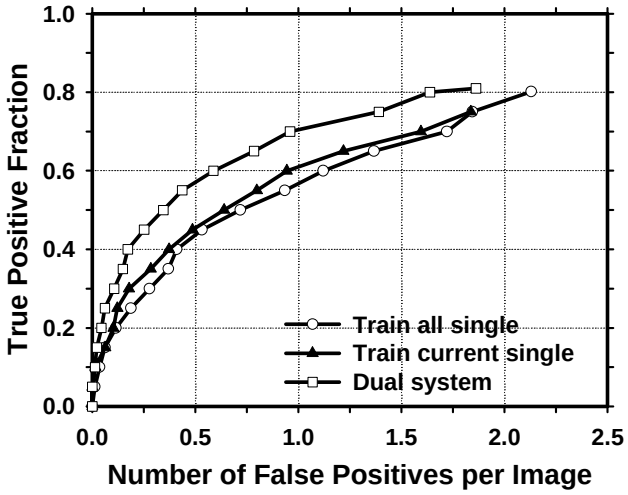
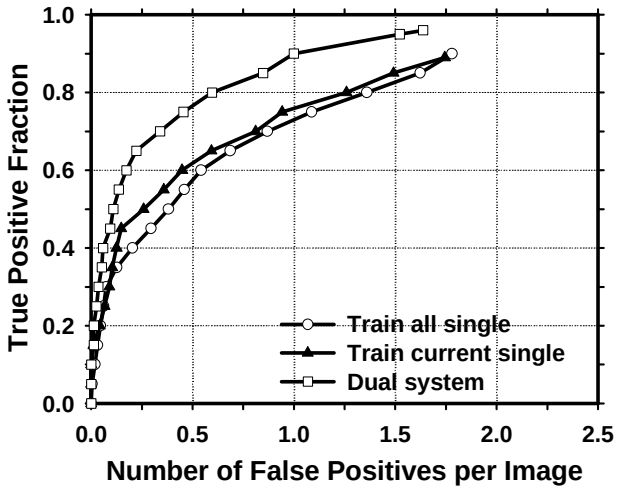


Figure 6. The test ROC curves for the BP-ANN classifiers from the two independent mass subsets. The ANN classifiers achieved an A_z value of 0.90±0.02 for test subset 1 and 0.89±0.01 for test subset 2 in the classification of mass and normal breast tissues.

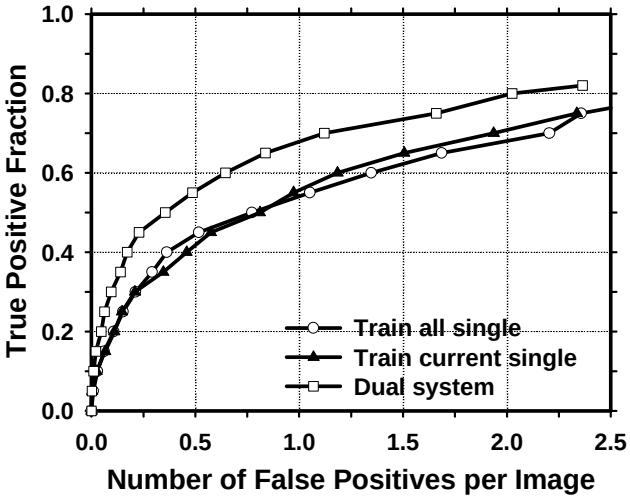


(a)

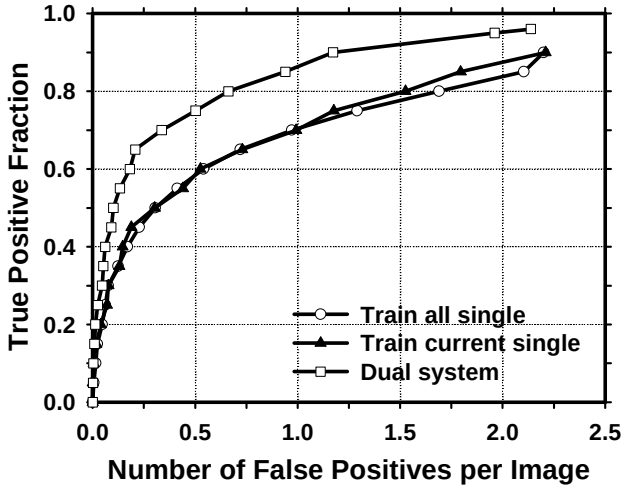


(b)

Figure 7. Comparison of the average test FROC curves obtained from averaging the FROC curves of the two independent average-mass subsets. Three CAD systems were compared: a single CAD system trained with average masses alone, a single CAD system trained with both the average and the subtle masses, and the dual CAD system. The FP rate was estimated from the mammograms with masses. (a) Image-based FROC curves, (b) Case-based FROC curves.

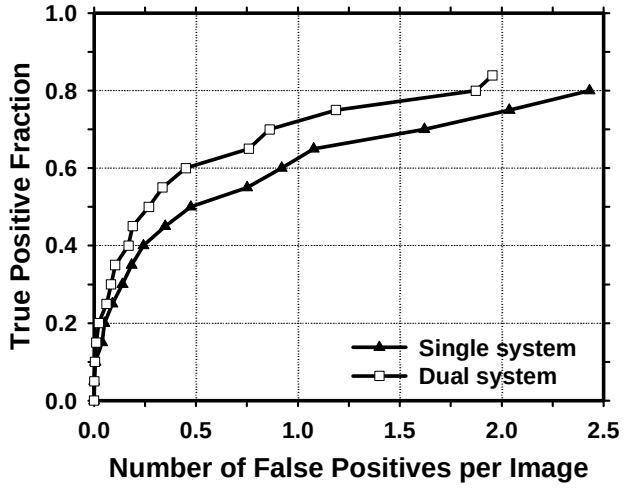


(a)

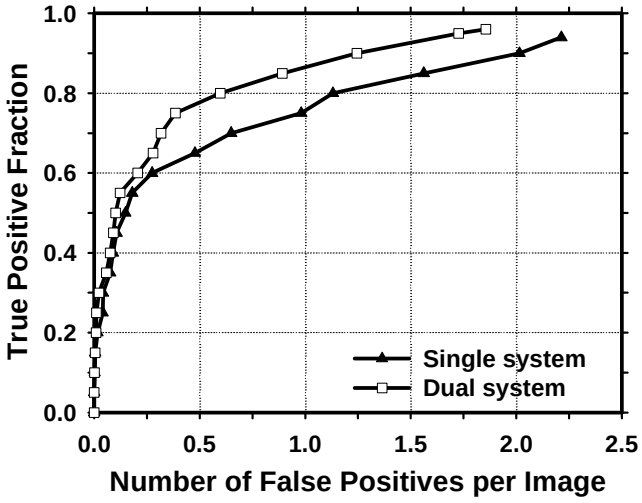


(b)

Figure 8. Comparison of the average test FROC curves obtained from averaging the FROC curves of the two independent average-mass subsets. Three CAD systems were compared: a single CAD system trained with average masses only, a single CAD system trained with the average and the subtle masses, and the dual CAD system. The FP rate was estimated from the mammograms without masses. (a) Image-based FROC curves, (b) Case-based FROC curves.

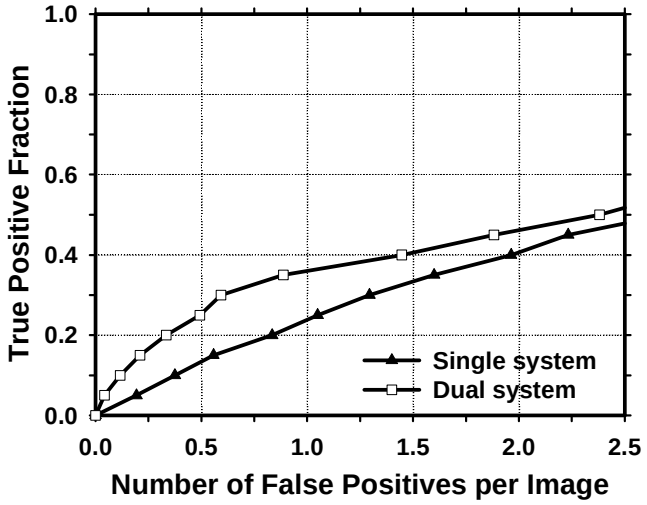


(a)

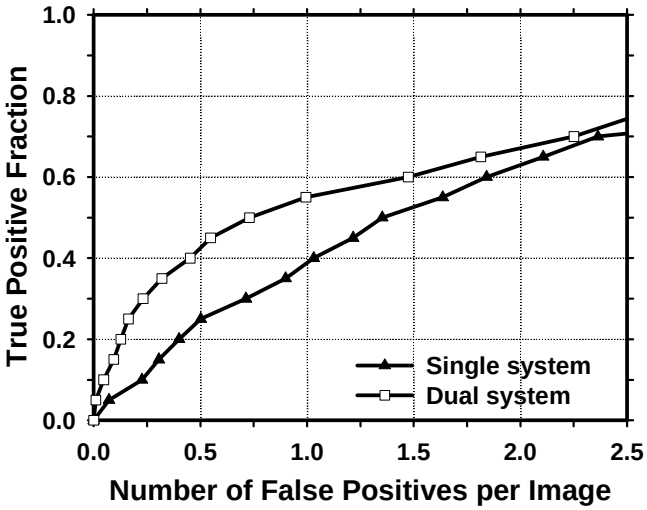


(b)

Figure 9. Comparison of the average test FROC curves of the single CAD system and the dual CAD system for detection of malignant masses in the average data set. The single system trained with average masses alone was used and the FP rate was estimated from the mammograms without masses. (a) Image-based FROC curves, (b) Case-based FROC curves.

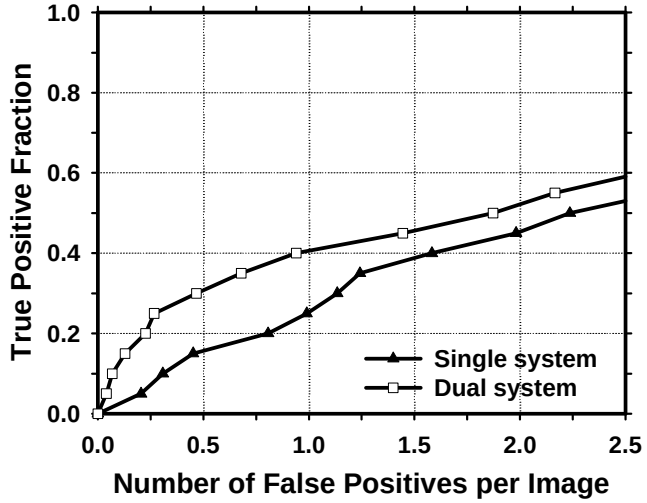


(a)

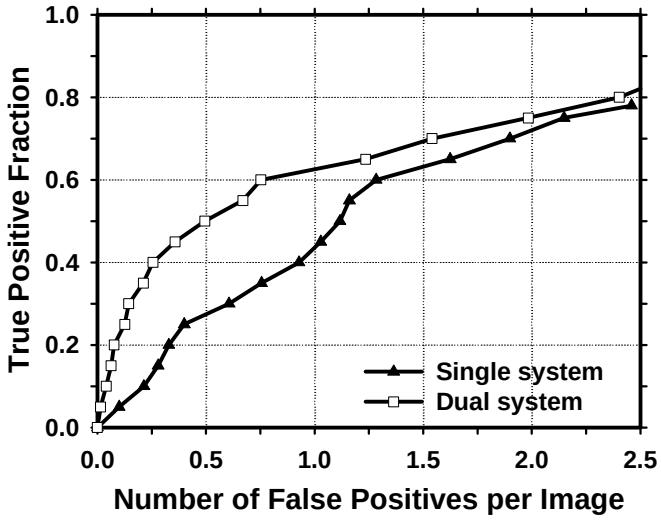


(b)

Figure 10. Comparison of the average test FROC curves for the single CAD system and the dual CAD system for detection of the subtle masses on the prior mammograms. The single CAD system trained with average masses alone was used and the FP rate was estimated from the mammograms without masses. (a) Image-based FROC curves, (b) Case-based FROC curves.



(a)



(b)

Figure 11. Comparison of the average test FROC curves for the single CAD system and the dual CAD system for detection of subtle malignant masses on the prior mammograms. The single CAD system trained with average masses alone was used and the FP rate was estimated from the mammograms without masses. (a) Image-based FROC curves, (b) Case-based FROC curves.

Regularized discriminant analysis for breast mass detection on full field digital mammograms

Jun Wei, Berkman Sahiner, Yiheng Zhang, Heang-Ping Chan, Lubomir M. Hadjiiski,
Chuan Zhou, Jun Ge, Yi-Ta Wu
Department of Radiology, University of Michigan, Ann Arbor

ABSTRACT

In computer-aided detection (CAD) applications, an important step is to design a classifier for the differentiation of the abnormal from the normal structures. We have previously developed a stepwise linear discriminant analysis (LDA) method with simplex optimization for this purpose. In this study, our goal was to investigate the performance of a regularized discriminant analysis (RDA) classifier in combination with a feature selection method for classification of the masses and normal tissues detected on full field digital mammograms (FFDM). The feature selection scheme combined a forward stepwise feature selection process and a backward stepwise feature elimination process to obtain the best feature subset. An RDA classifier and an LDA classifier in combination with this new feature selection method were compared to an LDA classifier with stepwise feature selection. A data set of 130 patients containing 260 mammograms with 130 biopsy-proven masses was used. All cases had two mammographic views. The true locations of the masses were identified by experienced radiologists. To evaluate the performance of the classifiers, we randomly divided the data set into two independent sets of approximately equal size for training and testing. The training and testing were performed using the 2-fold cross validation method. The detection performance of the CAD system was assessed by free response receiver operating characteristic (FROC) analysis. The average test FROC curve was obtained by averaging the FP rates at the same sensitivity along the two corresponding test FROC curves from the 2-fold cross validation. At the case-based sensitivities of 90%, 80% and 70% on the test set, our RDA classifier with the new feature selection scheme achieved an FP rate of 1.8, 1.1, and 0.6 FPs/image, respectively, compared to 2.1, 1.4, and 0.8 FPs/image with stepwise LDA with simplex optimization**.** Our results indicate that RDA in combination with the sequential forward inclusion-backward elimination feature selection method can improve the performance of mass detection on mammograms. Further work is underway to optimize the feature selection and classification scheme and to evaluate if this approach can be generalized to other CAD classification tasks.

Keywords: computer-aided detection, full field digital mammogram, mass detection, regularized discriminant analysis, feature selection

1. INTRODUCTION

Breast cancer is the most common cancer among American women[1]. Early detection and diagnosis can significantly increase the survival rate[2-4]. Recent clinical studies have shown that computer-aided detection (CAD) systems are helpful for increasing radiologists' accuracy in detecting breast cancers[5-8].

We have been developing CAD systems for detection and characterization of mammographic masses and microcalcifications. Detection of masses on mammograms is more challenging than detection of microcalcifications because the normal fibroglandular tissue in the breast causes false positives (FPs) by mimicking masses and causes false negatives due to overlapping with the lesions. Therefore, mass detection systems generally have lower sensitivity and higher FP rate than microcalcification detection systems. We are investigating methods to improve the overall performance of our CAD systems.

False positive (FP) classification is an important step in a CAD system. The basic approach in two-class classification is to assign an unknown sample to one of the two classes on the basis of a multidimensional feature space. A number of methods have been proposed in previous studies[9-11]. Most of the methods are based on linear discriminant analysis (LDA), artificial neural networks, and rule-based classifiers[12]. Recently, support vector machines were used to

Medical Imaging 2006: Image Processing, edited by Joseph M. Reinhardt, Josien P. W. Pluim,
Proc. of SPIE Vol. 6144, 61445P, (2006) · 0277-786X/06/$15 · doi: 10.1117/12.654897

classify the malignant and benign clustered microcalcifications on mammograms[13]. In medical imaging application, a main problem during the classifier design is the finite sample size available which biases the performance of the trained classifier for unknown cases. In this study, we are investigating the performance of a regularized discriminant analysis (RDA) classifier in combination with a feature selection method for classification of the masses and normal tissues detected on full field digital mammograms (FFDMs).

2. MATERIALS AND METHODS

2.1 Materials

IRB approval was obtained prior to the commencement of this investigation. The images used in this study were acquired at the University of Michigan with a GE Senographe 2000D FFDM system before biopsy. The GE system has a CsI phosphor/a:Si active matrix flat panel digital detector with a pixel size of $100\mu m \times 100\mu m$ and 14 bits per pixel. A data set of 130 cases was used. All cases had two mammographic views, the craniocaudal (CC) view and the mediolateral oblique (MLO) view or the lateral (LM or ML) view. The data set contained 130 biopsy-proven masses. The true locations of the masses were identified by a Mammography Quality Standards Act radiologist.

2.2 Methods

2.2.1 Discriminant Analysis

Assume that the class distributions are multivariate normal in a two-class classification problem. Under this condition, discriminant analysis models differ essentially by the specific assumptions on the mean vectors and covariance matrices of the group conditional densities. The most commonly used model is linear discriminant analysis (LDA) which assumes that the group conditional distributions are multivariate normal distributions with mean vectors μ_k, where k = 1, 2 is the class index, and equal covariance matrix Σ. The definition of LDA is given in Eq. (1).

$$Y = (\mu_1 - \mu_2)^T \Sigma^{-1} X \tag{1}$$

where $X^T=(x_1, \ldots, x_n)$ is the feature vector of a sample and n is the dimensionality of the feature space. If the covariance matrices are not equal, one can use quadratic discriminant analysis (QDA), which has a quadratic term for the feature vector in its model. The definition of QDA is described in Eq. (2).

$$Y = \frac{1}{2} X^T (\Sigma_2^{-1} - \Sigma_1^{-1}) X - (\mu_2^T \Sigma_2^{-1} - \mu_1^T \Sigma_1^{-1}) X \tag{2}$$

The parameters in LDA and QDA are usually unknown and have to be estimated from training samples. In medical imaging applications, the sample size may be very small in comparison with the dimensionality of the feature space. A regularization technique for discriminant analysis, referred to as regularized discriminant analysis (RDA)[14], makes use of a complexity parameter and a shrinkage parameter to design an intermediate classification model between LDA and QDA. The covariance matrices can thus be written as:

$$\hat{\Sigma}_k = (1-\gamma)\Sigma_k + \frac{\gamma}{p} tr[\Sigma_k] I \quad , \qquad k=1, 2 \tag{3}$$

where I is the identity matrix, γ and p are the complexity parameter and the shrinkage parameter, respectively. In this work, we investigated the use of the RDA classifier for FP reduction in a mass CAD system.

2.2.2 CAD System Overview

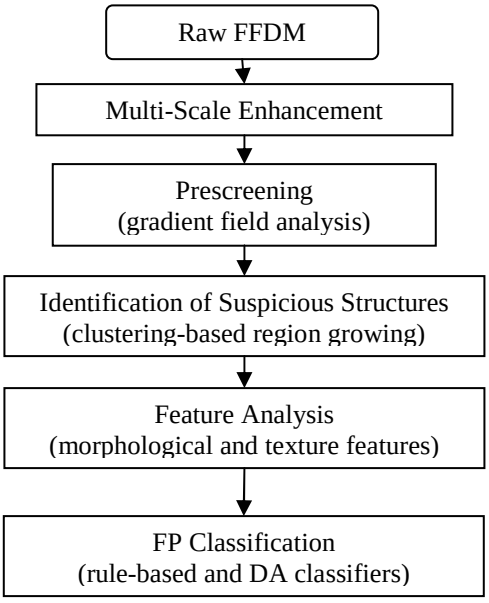
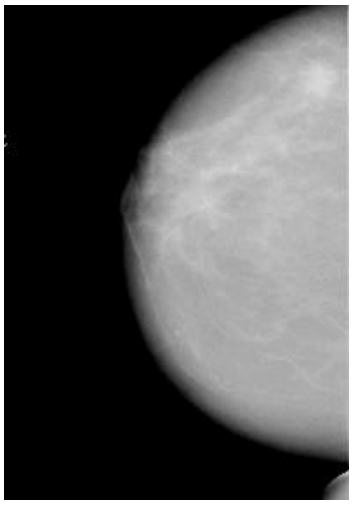
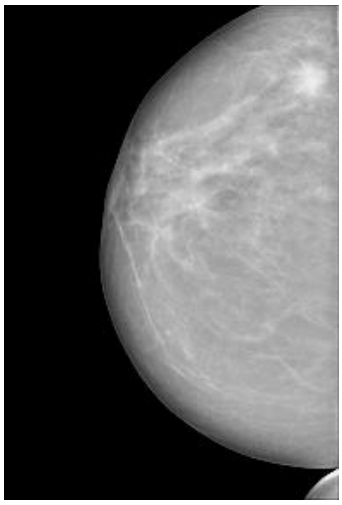
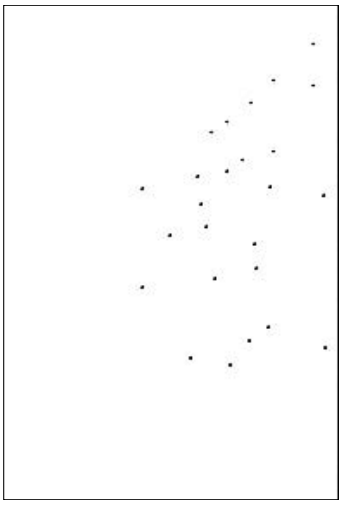
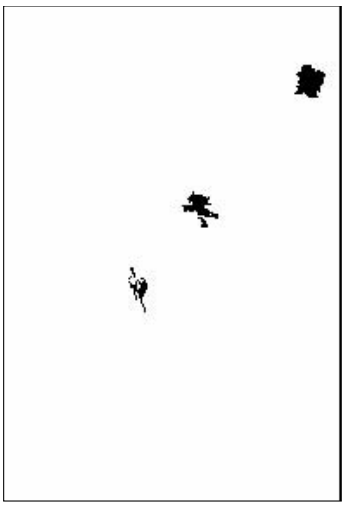
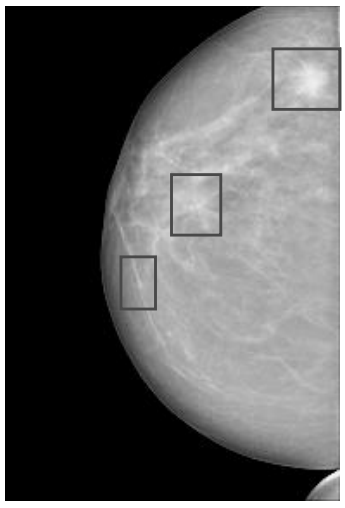


Figure 1. Block diagram of CAD system for mass detection on FFDMs.

Our CAD system consists of five processing steps: (1) preprocessing by using multi-scale enhancement, (2) pre-screening of mass candidates, (3) identification of suspicious objects, (4) feature extraction and analysis, and (5) FP reduction by classification of normal tissue structures and masses. The block diagram for the scheme is shown in Figure 1. FFDMs generally are pre-processed with proprietary methods before being displayed to readers. In an effort to develop a CAD system that is less dependent on specific FFDM systems, the raw digital images are used as input to our system. A preprocessing scheme based on a multi-resolution method[15] has been developed for image enhancement. This scheme consists of three steps. First, the boundary of the breast is detected automatically by using Otsu's method[16]. Second, the Laplacian pyramid is used to decompose the image into multi-scales. A nonlinear weight function is designed to enhance each high-pass component. Finally, the Gaussian pyramid is used to reconstruct the multi-scales. An example of an original mammogram and the enhanced mammogram are shown in Figs. 2(a) and 2(b), respectively. After preprocessing, gradient field analysis was used to detect the mass candidates from the preprocessed FFDMs. The suspicious objects are then identified by using a clustering based region growing method. Figures 2(c) and 2(d) show the initial detection locations and the grown objects, respectively. For each suspicious object, eleven morphologic features are extracted and rule-based and discriminant classifiers are trained to remove the detected normal structures that are substantially different from breast masses. Global and local multiresolution texture analysis[17, 18] are performed in each region of interest by using the spatial gray level dependence matrix. Finally, discriminant classification is used to identify potential breast masses. Further details of this algorithm can be found in the literature[19].

In order to obtain the best texture feature subset and reduce the dimensionality of the feature space to design an effective classifier, feature selection was applied to the training set. Stepwise LDA feature selection with Wilks' lambda as the selection criterion was employed in our previous study. Simplex optimization procedure was used to choose the best set of feature selection parameters which includes a threshold F_{in} for feature entry, a threshold F_{out} for feature removal, and a tolerance threshold T for excluding features that have high correlation with the features already in the selected pool. In this study, we compared a new stepwise feature selection procedure with the current method. In the proposed method, a feature selection scheme which combines forward stepwise feature selection and backward stepwise feature elimination is used to obtain the best feature subset, using the area under the receiver operating characteristic (ROC) curve, A_z, as the selection criterion instead of Wilks' lambda. We evaluated the classifier performance using a leave-one-case-out resampling scheme within the training set, the test discriminant scores from the left-out cases were analyzed using ROC methodology. The discriminant scores were input as the decision variable in

sensitivity along the two corresponding test FROC curves from the 2-fold cross validation. The CAD system using RDA with the new feature selection method achieved an image-based sensitivity of 60%, 65%, and 70% at 1.1, 1.4, and 1.6 FPs/image, respectively, compared with 1.4, 1.7, and 2.1 FPs/image for the CAD system using LDA with the new feature selection method. The CAD system with stepwise LDA and simplex optimization achieved FP rates of 1.6, 1.9, and 2.2 FPs/image, respectively, at the same sensitivities, which were comparable to the FP rates of the CAD system using LDA with the new feature selection method. For case-based FROC analysis, the results are summarized in Table 1. Figures 3 and 4 show the comparison of the image-based and case-based average FROC curves of the CAD systems using the three different classification methods, respectively.

Table 1. Comparison of case-based performance of three methods. OFS: stepwise feature selection with simplex optimization. NFS: feature selection combining forward feature selection and backward feature elimination.

TP	FPs/image		
	LDA-OFS	LDA-NFS	RDA-NFS
70%	0.8	0.7	0.6
80%	1.4	1.3	1.1
90%	2.1	2.2	1.8

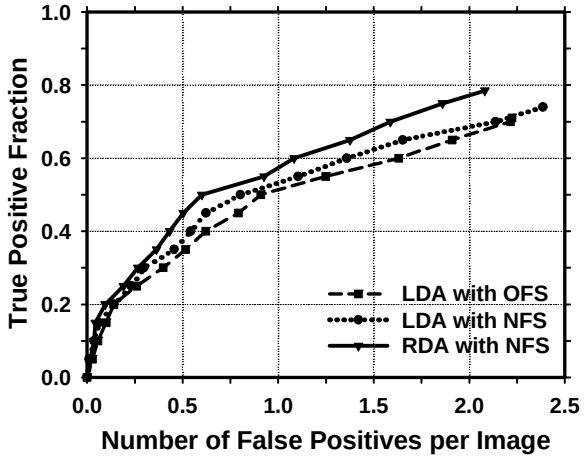


Figure 3. Comparison of image-based FROC curves. OFS: stepwise feature selection with simplex optimization. NFS: feature selection combining forward feature selection and backward feature elimination.

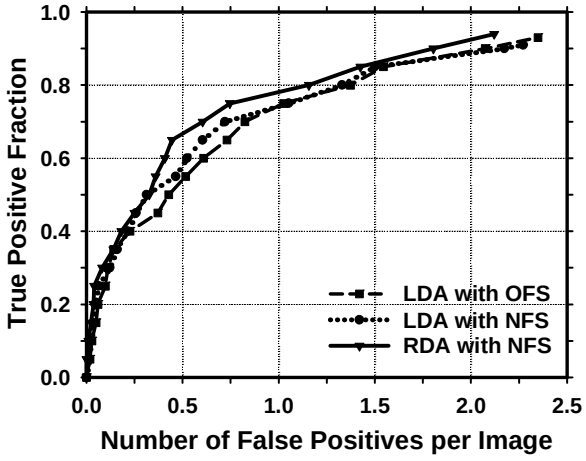


Figure 4. Comparison of case-based FROC curves. OFS: stepwise feature selection with simplex optimization. NFS: feature selection combining forward feature selection and backward feature elimination.

4. DISCUSSION AND CONCLUSIONS

We previously developed a CAD system for detection of masses on FFDMs. In this study, we investigated the use of an RDA classifier with a new feature selection method. Our results indicated that the new FP classifier can improve the overall performance of our CAD system. Further work is underway to optimize the feature selection and classification scheme and to evaluate if this approach can be generalized to other CAD classification tasks.

ACKNOWLEGMENTS

This work is supported by U. S. Army Medical Research and Materiel Command grants DAMD 17-02-1-0214 and W81XWH-04-1-0475 and USPHS grant CA95153. The content of this paper does not necessarily reflect the

position of the government and no official endorsement of any equipment and product of any companies mentioned should be inferred. The authors are grateful to Charles E. Metz, Ph.D., for the LABROC program.

REFERENCES

1. "American cancer society, www.Cancer.Org 2004, "Statistics for 2004"," American Cancer Society (2004).
2. C. R. Smart, R. E. Hendrick, J. H. Rutledge and R. A. Smith, "Benefit of mammography screening in women ages 40 to 49 years: Current evidence from randomized controlled trials," *Cancer* **75**, 1619-1626, 1995.
3. S. A. Feig, C. J. D'orsi, R. E. Hendrick, V. P. Jackson, D. B. Kopans, B. Monsees, E. A. Sickles, C. B. Stelling, M. Zinninger and P. Wilcox-Buchalla, "American college of radiology guidelines for breast cancer screening," *AJR Am J Roentgenol.* **171**, 29-33, 1998.
4. B. Cady and J. S. Michaelson, "The life-sparing potential of mammographic screening," *CANCER* **91**, 1699-1703, 2001.
5. T. W. Freer and M. J. Ulissey, "Screening mammography with computer-aided detection: Prospective study of 12,860 patients in a community breast center," *Radiology* **220**, 781-786, 2001.
6. S. V. Destounis, P. Dinitto, W. Logan-Young, E. Bonaccio, M. L. Zuley and K. M. Willison, "Can computer-aided detection with double reading of screening mammograms help decrease the false-negative rate? Initial experience," *Radiology* **232**, 578-584, 2004.
7. M. A. Helvie, L. M. Hadjiiski, E. Makariou, H. P. Chan, N. Petrick, B. Sahiner, S. C. B. Lo, M. Freedman, D. Adler, J. Bailey, C. Blane, D. Hoff, K. Hunt, L. Joynt, K. Klein, C. Paramagul, S. Patterson and M. A. Roubidoux, "Sensitivity of noncommercial computer-aided detection system for mammographic breast cancer detection - a pilot clinical trial," *Radiology* **231**, 208-214, 2004.
8. R. L. Birdwell, P. Bandodkar and D. M. Ikeda, "Computer-aided detection with screening mammography in a university hospital setting," *Radiology* **236**, 451-457, 2005.
9. H. P. Chan, S. C. B. Lo, B. Sahiner, K. L. Lam and M. A. Helvie, "Computer-aided detection of mammographic microcalcifications: Pattern recognition with an artificial neural network," *Medical Physics* **22**, 1555-1567, 1995.
10. H. P. Chan, B. Sahiner, R. F. Wagner and N. Petrick, "Classifier design for computer-aided diagnosis: Effects of finite sample size on the mean performance of classical and neural network classifiers," *Medical Physics* **26**, 2654-2668, 1999.
11. B. Sahiner, H. P. Chan, N. Petrick, D. Wei, M. A. Helvie, D. D. Adler and M. M. Goodsitt, "Classification of mass and normal breast tissue: A convolution neural network classifier with spatial domain and texture images," *IEEE Transactions on Medical Imaging* **15**, 598-610, 1996.
12. Q. Li and K. Doi, "Analysis and minimization of overtraining effect in rule-based classifiers for computer-aided diagnosis," *Med. Phys* **33**, 320–328, 2006.
13. J. Wei, B. Sahiner, L. M. Hadjiiski, H.-P. Chan, N. Petrick, M. A. Helvie, M. A. Roubidoux, J. Ge and C. Zhou, "Computer aided detection of breast masses on full field digital mammograms," *Med.Phys* 2005 (Submitted).
14. J. Friedman, "Regularized discriminant analysis," *Journal of the American Statistical Association* **84**, 165-175, 1989.
15. P. J. Burt and E. H. Adelson, "The laplacian pyramid as a compact image code," *IEEE Transactions on Communications* **COM-31**, 337-345, 1983.
16. N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. System, Man, Cybernetics* **9**, 62-66, 1979.
17. D. Wei, H. P. Chan, M. A. Helvie, B. Sahiner, N. Petrick, D. D. Adler and M. M. Goodsitt, "Classification of mass and normal breast tissue on digital mammograpms: Multiresolution texture analysis," *Medical Physics* **22**, 1501-1513, 1995.
18. D. Wei, H. P. Chan, N. Petrick, B. Sahiner, M. A. Helvie, D. D. Adler and M. M. Goodsitt, "False-positive reduction technique for detection of masses on digital mammograms: Global and local multiresolution texture analysis," *Medical Physics* **24**, 903-914, 1997.
19. J. Wei, B. Sahiner, L. M. Hadjiiski, H. P. Chan, N. Petrick, M. A. Helvie, M. A. Roubidoux, J. Ge and C. Zhou, "Computer aided detection of breast masses on full field digital mammograms," *Medical Physics* **32**, 2827-2838, 2005.

Two-view information fusion for improvement of computer-aided detection (CAD) of breast masses on mammograms

Jun Wei, Berkman Sahiner, Lubomir M. Hadjiiski, Heang-Ping Chan, Mark A. Helvie,
Marilyn A. Roubidoux, Chuan Zhou, Jun Ge, Yiheng Zhang
Department of Radiology, The University of Michigan, Ann Arbor, MI 48109

ABSTRACT

We are developing a two-view information fusion method to improve the performance of our CAD system for mass detection. Mass candidates on each mammogram were first detected with our single-view CAD system. Potential object pairs on the two-view mammograms were then identified by using the distance between the object and the nipple. Morphological features, Hessian feature, correlation coefficients between the two paired objects and texture features were used as input to train a similarity classifier that estimated a similarity scores for each pair. Finally, a linear discriminant analysis (LDA) classifier was used to fuse the score from the single-view CAD system and the similarity score. A data set of 475 patients containing 972 mammograms with 475 biopsy-proven masses was used to train and test the CAD system. All cases contained the CC view and the MLO or LM view. We randomly divided the data set into two independent sets of 243 cases and 232 cases. The training and testing were performed using the 2-fold cross validation method. The detection performance of the CAD system was assessed by free response receiver operating characteristic (FROC) analysis. The average test FROC curve was obtained from averaging the FP rates at the same sensitivity along the two corresponding test FROC curves from the 2-fold cross validation. At the case-based sensitivities of 90%, 85% and 80% on the test set, the single-view CAD system achieved an FP rate of 2.0, 1.5, and 1.2 FPs/image, respectively. With the two-view fusion system, the FP rates were reduced to 1.7, 1.3, and 1.0 FPs/image, respectively, at the corresponding sensitivities. The improvement was found to be statistically significant ($p<0.05$) by the AFROC method. Our results indicate that the two-view fusion scheme can improve the performance of mass detection on mammograms.

Keywords: computer-aided detection, two-view fusion, mass detection, AFROC analysis

1. INTRODUCTION

Breast cancer is one of the leading causes of cancer mortality among women[1]. There is considerable evidence that early diagnosis and treatment significantly improves the chance of survival for patients with breast cancer [2-5]. Although mammography has a high sensitivity for detection of breast cancers when compared to other imaging modalities, studies indicate that radiologists do not detect all carcinomas that are visible upon retrospective analyses of the images[6-11]. It has been shown that computer-aided detection (CAD) can improve the sensitivity of mammography in prospective clinical trials[12-15]. CAD is thus a viable cost-effective alternative to double reading by radiologists.

The mass detection systems to-date generally employed a single-view detection approach using various techniques for prescreening of mass candidates and classification of true and false positives[16-25]. We have been developing CAD systems for detection of mammographic masses on full field digital mammograms (FFDMs)[25] and screening film mammograms (SFMs)[22]. Our previous study[23] showed that two-view fusion method can improve the performance of a CAD system for mass detection on mammograms. In this study, our purpose is to improve the performance of the two-view information fusion method and to test our method in a relatively larger data set.

2. MATERIALS AND METHODS

2.1 Materials

Medical Imaging 2006: Image Processing, edited by Joseph M. Reinhardt, Josien P. W. Pluim,
Proc. of SPIE Vol. 6144, 614424, (2006) · 0277-786X/06/$15 · doi: 10.1117/12.654593

All mammograms in this study were collected from patient files in the Department of Radiology at the University of Michigan with Institutional Review Board (IRB) approval. The mammograms were digitized with a LUMISYS 85 laser film scanner with a pixel size of 50μm×50μm and 4096 gray levels. The scanner was calibrated to have a linear relationship between gray levels and optical densities (O.D.) from 0.1 to greater than 3 O.D. units. The nominal O.D. range of the scanner is 0–4. The full resolution mammograms were first smoothed with a 2×2 box filter and subsampled by a factor of 2, resulting in images with a pixel size of 100μm×100μm. These images were used for the input of our CAD system. The data set we used in this study contained 475 cases, of which 464 cases had the two-view mammograms (the craniocaudal (CC) view and the mediolateral oblique (MLO) view or the lateral view) and 11 cases had four-view mammograms, resulting in a total of 972 mammograms. All mammograms were obtained before biopsy. There were 475 biopsy-proven masses in this data set.

2.2 Methods

2.2.1 Single-view System Overview

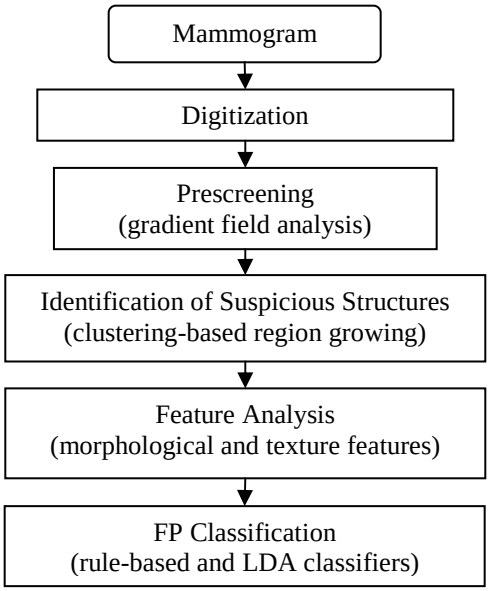
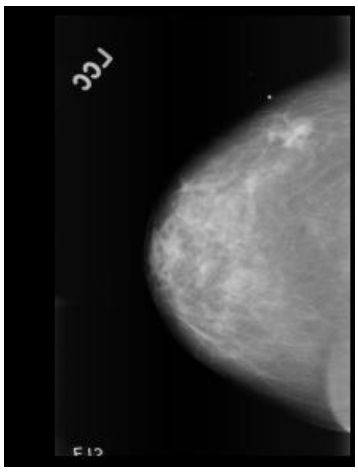
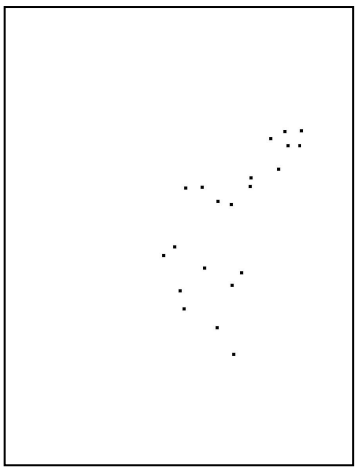
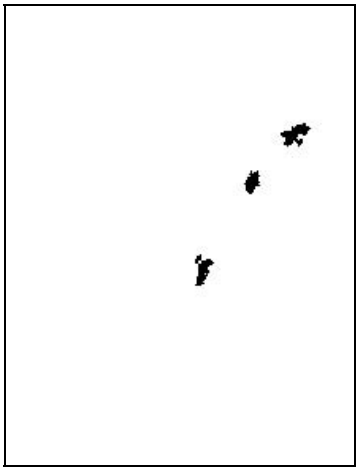


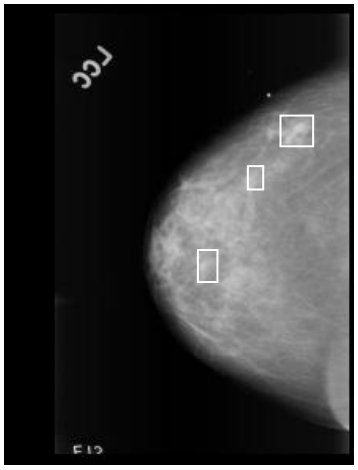
Figure 1. Block diagram of a single CAD system for mass detection on mammograms.

Our single-view CAD system consists of five processing steps: 1) pre-screening of mass candidates, 2) identification of suspicious objects, 3) extraction of morphological and texture features, and 4) classification between the normal and the abnormal regions by using rule-based and LDA classifiers. The block diagram for the single-view CAD system is shown in Figure 1. Figure 2 shows an example demonstrating the processing steps with our computer-aided mass detection system. For the pre-screening stage, we have developed a two-stage gradient field analysis method which combines the shape information of masses on mammograms with the gray level information of the local object segmented by a region growing technique in the second stage to refine the gradient field analysis. The gradient field analysis is used to determine locations of high convergence of radial gradient in the image. A region of interest (ROI) is then identified with its center placed at each location of high gradient convergence. The object in each ROI is segmented by a region growing method in which the location of high gradient convergence is used as the starting point. Figures 2(b) and 2(c) show the initial detection locations and the grown objects, respectively. After region growing, all connected pixels constituting the object are labeled. Finally, the gradient convergence at the center location of the ROI is recalculated within the segmented object. The objects whose new gradient convergence is lower than 80% of the original value are rejected. After prescreening, the suspicious objects are identified by using a clustering-based region growing method. For each suspicious object, eleven morphological features are extracted. Rule-based and LDA classifiers are trained to remove the detected normal structures that are substantially different from breast masses. Global and local multiresolution texture analyses are performed in each ROI by using the spatial gray level dependence









object. Manually identified nipple locations are used for the registration in this study. We are developing an automated nipple detection technique[26] and the automated method will be used when it reaches high accuracy. Similarity measures between each pair of objects are derived from the pairs of individual object features. The similarity features include morphological features, Hessian feature, correlation coefficients between the two paired objects and texture features. A similarity classifier is trained to distinguish between true and false pairs by merging the similarity features into a similarity score for each object. The similarity score and the single-view object score of the object are then fused to form a final score for the object. Our two-view system is summarized in Figure 3.

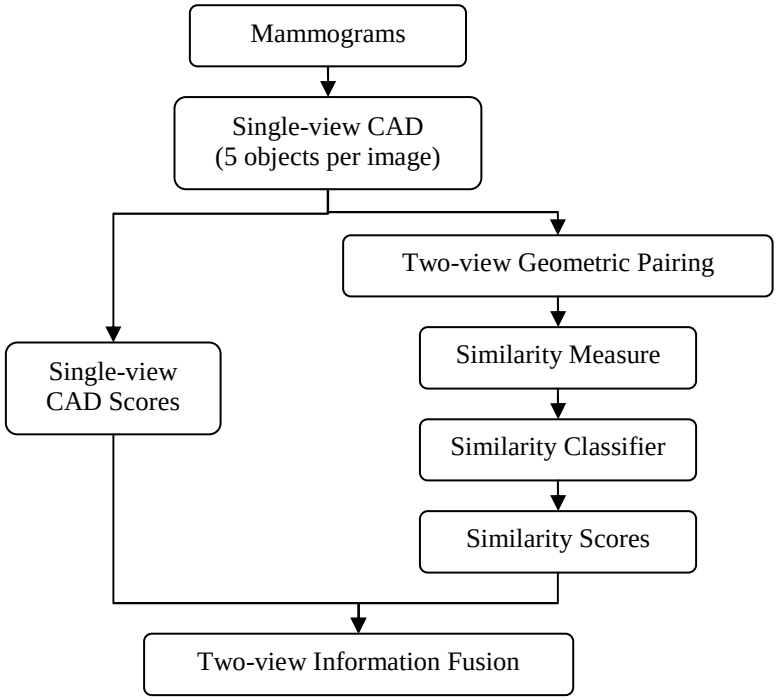


Figure 3. Block diagram of the two-view CAD system for mass detection on mammograms.

3. Experimental Results

We randomly separated the cases in our data set into two independent equal sized data sets: 243 cases with 494 images and 232 cases with 478 images. The training and testing were performed using the 2-fold cross validation method. The detection performance of the CAD system was assessed by free response receiver operating characteristic (FROC) analysis. FROC curves were presented on a per-mammogram and a per-case basis. For mammogram-based FROC analysis, the mass on each mammogram was considered an independent true object. For case-based FROC analysis, the same mass imaged on the two-view mammograms was considered to be one true object and the detection of either or both masses on the two views was considered to be a true-positive (TP). To evaluate the overall test performance, an average test FROC curve was obtained from averaging the FP rates at the same sensitivity along the two corresponding test FROC curves from the 2-fold cross validation. When the single-view CAD system was applied to the test set, the FPs/image were 2.0, 1.5, and 1.2 at the case-based sensitivities of 90%, 85% and 80%, respectively. With the two-view CAD system, the FP rates were improved to 1.7, 1.3, and 1.0 FPs/image at the same case-based sensitivities. Figure 4 and 5 shows the comparison of the test performance of the single-view CAD system and the two-view CAD systems by using image-based and case-based average FROC curves, respectively. To analyze the improvement in the FROC curves statistically, an alternative free-response ROC (AFROC) method[27] was employed. In the AFROC method, false-positive images (FPI) instead of FPs per image are counted. The confidence rating of an FPI is determined by the highest confidence FP decision on the image regardless of how many lower confidence FP decisions

are made on the same image. The ROCKIT software developed by Metz et al[28] is used to analyze the AFROC data. The comparison of the A_1 and the p values is summarized in Table 1.

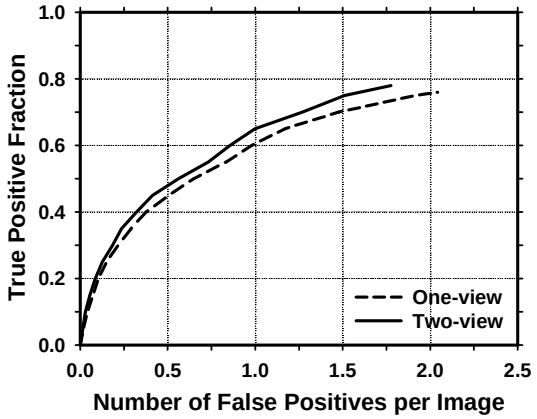


Figure 4. Image-based average FROC curves obtained from averaging the corresponding FROC curves of the two test subsets. Single-view: detection by the single-view CAD system. Two-view: detection by the two-view CAD system.

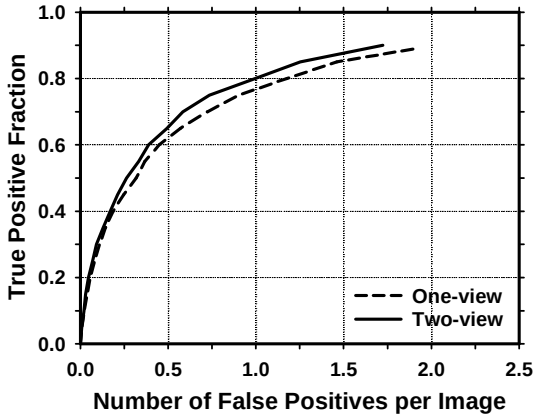


Figure 5. Case-based average FROC curves obtained from averaging the corresponding FROC curves of the two test subsets. Single-view: detection by the single-view CAD system. Two-view: detection by the two-view CAD system.

Table 1. Estimation of the statistical significance in the difference between the FROC performances of the single-view CAD system and the two-view CAD system.

	A_1 (AFROC)	
	Test Set 1	Test Set 2
One-view CAD	0.52	0.51
Two-view CAD	0.55	0.54
P Value	<0.0001	<0.0001

4. DISCUSSION AND CONCLUSIONS

In this study, we developed a two-view CAD system to improve the computerized detection of masses on mammograms. The two-view CAD system is different from case-based scoring, in which detection of the same mass in either the CC view or the MLO view will be counted as a true positive, in that the detected objects in the two views are correlated and analyzed for similarity and the likelihood score of a mass detected in both views may be enhanced compared with FPs. Our results indicate that two-view fusion can significantly improve the overall performance of the single-view CAD system. Future work will include automated identification of nipple locations and optimization of the fusion scheme in our system.

ACKNOWLEDGMENTS

This work is supported by USPHS grant CA95153, U. S. Army Medical Research and Materiel Command grants DAMD 17-02-1-0214 and W81XWH-04-1-0475. The content of this paper does not necessarily reflect the position of the government and no official endorsement of any equipment and product of any companies mentioned should be inferred. The authors are grateful to Charles E. Metz, Ph.D., for the LABROC and ROCKIT programs.

REFERENCES

1. "American cancer society, www.Cancer.Org 2004, "Statistics for 2004"," American Cancer Society (2004).
2. C. R. Smart, R. E. Hendrick, J. H. Rutledge and R. A. Smith, "Benefit of mammography screening in women ages 40 to 49 years: Current evidence from randomized controlled trials," *Cancer* **75**, 1619-1626, 1995.
3. C. Byrne, C. R. Smart, C. Cherk and W. H. Hartmann, "Survival advantage differences by age: Evaluation of the extended follow-up of the breast cancer detection demonstration project," *Cancer* **74**, 301-310, 1994.
4. S. A. Feig and R. E. Hendrick, "Risk, benefit, and controversies in mammographic screening," *Syllabus: A categorical course in physics technical aspects of breast imaging*, A. G. Haus and M. J. Yaffe, Eds., 119-135, Radiological Society of North America, Inc, 1993.
5. H. Seidman, S. K. Gelb, E. Silverberg, N. Laverda and J. A. Lubera, "Survival experience in the breast cancer detection demonstration project," *CA Cancer J Clin.* **37**, 258-290, 1987.
6. B. J. Hillman, L. L. Fajardo, T. B. Hunter, B. Mockbee, C. E. Cook, R. M. Hagaman, J. C. Bjelland, C. S. Frey and C. J. Harris, "Mammogram interpretation by physician assistants," *Am J Roentgenology* **149**, 907-911, 1987.
7. L. W. Bassett, D. H. Bunnell, R. Jahanshahi, R. H. Gold, R. D. Arndt and J. Linsman, "Breast cancer detection: One versus two views," *Radiology* **165**, 95-97, 1987.
8. M. G. Wallis, M. T. Walsh and J. R. Lee, "A review of false negative mammography in a symptomatic population," *Clinical Radiology* **44**, 13-15, 1991.
9. J. A. Harvey, L. L. Fajardo and C. A. Innis, "Previous mammograms in patients with impalpable breast carcinomas: Retrospective vs blinded interpretation," *American Journal of Roentgenology* **161**, 1167-1172, 1993.
10. R. E. Bird, T. W. Wallace and B. C. Yankaskas, "Analysis of cancers missed at screening mammography," *Radiology* **184**, 613-617, 1992.
11. C. Beam, P. Layde and D. Sullivan, "Variability in the interpretation of screening mammograms by us radiologists," *Arch Intern Med* **156**, 209-213, 1996.
12. T. W. Freer and M. J. Ulissey, "Screening mammography with computer-aided detection: Prospective study of 12,860 patients in a community breast center," *Radiology* **220**, 781-786, 2001.
13. M. A. Helvie, L. M. Hadjiiski, E. Makariou, H. P. Chan, N. Petrick, B. Sahiner, S. C. B. Lo, M. Freedman, D. Adler, J. Bailey, C. Blane, D. Hoff, K. Hunt, L. Joynt, K. Klein, C. Paramagul, S. Patterson and M. A. Roubidoux, "Sensitivity of noncommercial computer-aided detection system for mammographic breast cancer detection - a pilot clinical trial," *Radiology* **231**, 208-214, 2004.
14. T. E. Cupples, "Impact of computer-aided detection (CAD) in a regional screening mammography program," *Radiology* **221(P)**, 520, 2001.
15. R. L. Birdwell, P. Bandodkar and D. M. Ikeda, "Computer-aided detection with screening mammography in a university hospital setting," *Radiology* **236**, 451-457, 2005.
16. S. L. Ng and W. F. Bischof, "Automated detection and classification of breast tumors," *Computers and Biomedical Research* **25**, 218-237, 1992.
17. N. Petrick, H. P. Chan, D. Wei, B. Sahiner, M. A. Helvie and D. D. Adler, "Automated detection of breast masses on mammograms using adaptive contrast enhancement and texture classification," *Medical Physics* **23**, 1685-1696, 1996.
18. B. Zheng, Y. H. Chang and D. Gur, "Computerized detection of masses in digitized mammograms using single-image segmentation and a multilayer topographic feature analysis," *Academic Radiology* **2**, 959-966, 1995.
19. N. Karssemeijer and G. Te Brake, "Detection of stellate distortions in mammograms," *IEEE Transactions on Medical Imaging* **15**, 611-619, 1996.
20. H. Kobatake and Y. Yoshinaga, "Detection of spicules on mammogram based on skeleton analysis," *IEEE Trans Med Img* **15**, 235-245, 1996.
21. W. Qian, L. H. Li and L. P. Clarke, "Image feature extraction for mass detection in digital mammography: Influence of wavelet analysis," *Medical Physics* **26**, 402-408, 1999.
22. N. Petrick, H. P. Chan, B. Sahiner, M. A. Helvie, S. Paquerault and L. M. Hadjiiski, "Breast cancer detection: Evalution of a mass detection algorithm for computer-aided diagnosis: Experience in 263 patients.," *Radiology* **224**, 217-224, 2002.

23. S. Paquerault, N. Petrick, H. P. Chan, B. Sahiner and M. A. Helvie, "Improvement of computerized mass detection on mammograms: Fusion of two-view information," *Medical Physics* **29**, 238-247, 2002.
24. B. Zheng, W. F. Good, D. R. Armfield, C. Cohen, T. Hertzberg, J. H. Sumkin and D. Gur, "Performance change of mammographic CAD schemes optimized with most-recent and prior image databases," *Academic Radiology* **10**, 283-288, 2003.
25. J. Wei, B. Sahiner, L. M. Hadjiiski, H. P. Chan, N. Petrick, M. A. Helvie, M. A. Roubidoux, J. Ge and C. Zhou, "Computer aided detection of breast masses on full field digital mammograms," *Medical Physics* **32**, 2827-2838, 2005.
26. C. Zhou, H. P. Chan, C. Paramagul, M. A. Roubidoux, B. Sahiner, L. M. Hadjiiski and N. Petrick, "Computerized nipple identification for multiple image analysis in computer-aided diagnosis," *Medical Physics* **31**, 2871-2882, 2004.
27. D. P. Chakraborty, "Maximum likelihood analysis of free-response receiver operating characteristic (FROC) data," *Med Phys* **16**, 561-568, 1989.
28. C. E. Metz, "Roc methodology in radiologic imaging," *Investigative Radiology* **21**, 720-733, 1986.